DOCUMENT RESUME

ED 139 443 IR 004 904

AUTHOR	Moore, Gwendolyn B.; And Others
TITLE	Accessing Individual Records from Personal Data Files Using Non-Unique Identifiers. Final Report. Computer Science & Technology Series.
INSTITUTION	National Bureau of Standards (DOC), Washington, D.C. Inst. for Computer Sciences and Technology.
REPORT NO	NBS-SP-500-2
PUB DATE	Feb 77
NOTE	315p.; For related documents, see IR 004 900-904; Best copy available
AVAILABLE FROM	Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 (C13.10:500-2, Stock No. 003-003-01726-4, $2.65)
EDRS PRICE	MF-$0.83 HC-$16.73 Plus Postage.
DESCRIPTORS	*Confidential Records; *Identification; *Information Retrieval; Probability; *Relevance (Information Retrieval); Reliability; *Search Strategies; State of the Art Reviews
IDENTIFIERS	Privacy Act of 1974

ABSTRACT

The Privacy Act of 1974 places restrictions on the Federal, state and local agencies' use of the Social Security account number as an identifier. For some agencies, compliance will involve changes in implementation of retrieval algorithms. This report describes methodology applicable to these changes in the more general context of the problem of retrieving individual records from files using non-unique identifiers. State-of-the-art retrieval techniques are discussed, a method for assigning reliability weights to various personal data elements is presented, file validation techniques for the error and omission rates of data items are suggested, and a retrieval probability model--designed to show likelihood of retrieval of a subject's record given a variety of populations, combinations of identifiers, and error/omission rates--is described. A methodology is developed for forming confidence factors from the established error/omission rates for combinations of non-unique identifiers that are candidates for use as retrieval keys. Use of these confidence factors as indices into the precision tables produced by the probability model is described. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED139443

COMPUTER SCIENCE & TECHNOLOGY:

Accessing Individual Records From Personal Data Files Using Non-Unique Identifiers

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Gwendolyn B. Moore
John L. Kuhns
Jeffrey L. Trefftzs
Christine A. Montgomery

Operating Systems, Inc.
21031 Ventura Boulevard
Woodland Hills, California 91364

Prepared for the
Institute for Computer Sciences and Technology
National Bureau of Standards
Washington, D.C. 20234

IR004904

U.S. DEPARTMENT OF COMMERCE

Dr. Betsy Ancker-Johnson, Assistant Secretary for Science and Technology

NATIONAL BUREAU OF STANDARDS, Ernest Ambler, Acting Director

Issued February 1977

Reports on Computer Science and Technology

The National Bureau of Standards has a special responsibility within the Federal Government for computer science and technology activities. The programs of the NBS Institute for Computer Sciences and Technology are designed to provide ADP standards, guidelines, and technical advisory services to improve the effectiveness of computer utilization in the Federal sector, and to perform appropriate research and development efforts as foundation for such activities and programs. This publication series will report these NBS efforts to the Federal computer community as well as to interested specialists in the academic and private sectors. Those wishing to receive notices of publications in this series should complete and return the form at the end of this publication.

National Bureau of Standards Special Publication 500-2
Nat. Bur. Stand. (U.S.), Spec. Publ. 500-2, 203 pages (Feb. 1977)
CODEN: XNBSAV

Library of Congress Cataloging in Publication Data

Main entry under title:

Accessing individual records from personal data files using non-unique identifiers.

(Computer science & technology) (NBS special publication ; 500-2)
Supt. of Docs. no.: C13.10:500-2

1. Personnel records–Data processing. 2. Civil service–United States–Personnel management. I. Moore, Gwendolyn B. II. Institute for Computer Sciences and Technology. III. Series. IV. Series: United States. National Bureau of Standards. Special publication ; 500-2.
QC100.U57 no. 500-2 [JK766.5] 602'.1s [353.001] 76-57950

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1977

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402
(Order by SD Catalog No. C13.10:500-2). Stock No. 003-003-01726-4 Price $2.65
(Add 25 percent additional for other than U.S. mailing).

FOREWORD

How does a Federal agency locate a personal record in one of its massive files with only a limited amount of information for guidance? The Privacy Act of 1974 presented numerous agencies with just this problem when it specified: "It shall be unlawful for any Federal, state, or local government agency to deny to any individual any right, benefit, or privilege provided by law because of such individual's refusal to disclose his Social Security account number."

Those Federal agencies denied the use of Social Security numbers are not the only ones interested in the solution to this problem. For instance, one might be required to locate individuals in a file who have been inadvertently or illegally recorded under more than one account number. Similarly, the results of a study of methodologies for retrieving an individual's record without the use of a universal identifier can be used to determine the threats to an individual's privacy.

This report was written in response to this growing need for retrieving information using non-unique identifiers. Presented are selected methodologies for assisting Federal agencies in selecting retrieval algorithms and name lookup techniques; in analyzing their data by the identification of weighting factors and statistical sampling for determining error and omission rates; and lastly, predicting the accuracy and efficiency of candidate retrieval keys.

Seymour Jeffery, Chief
Systems and Software Division
Institute for Computer Sciences
 and Technology
National Bureau of Standards

ACKNOWLEDGMENTS

The authors would like to express their appreciation to all those who contributed to the preparation of this document through their willingness to discuss their relevant experience and expertise. Particular thanks are due two members of the NBS Institute for Computer Sciences and Technology: Dr. Thomas C. Lowe, who coordinated this effort from its inception, gave incentive and direction to the project while assisting in every way possible with preparation of the final report; and John L. Berg, whose comments and suggestions have been invaluable throughout the course of the project.

TABLE OF CONTENTS

Page

1.0 Introduction. 1

1.1 The Legislation. 1
1.2 Retrieval of Individual Records. 3
1.3 Statistical vs. Dossier Files. 4
1.4 Privacy vs. Security 6
1.5 Summary. 8

2.0 State-of-the-Art Retrieval Techniques 9

2.1 Introduction 9
2.2 Varying Selective Values of Non-Unique Identifiers . . . 10
2.3 Importance of the Name Variable in Retrieval Through Use of Non-Unique Identifiers. 12
2.4 Selected State-of-the-Art for Name Lookup Techniques . . 12
2.5 Implementation of Name Lookup Techniques 26
2.6 Other Search Strategies. 29

3.0 Identification of Weighting Factors 32

3.1 Accuracy Constraints Imposed by the Privacy Act. 32
3.2 Determination of Weighting Factors for Existing Files. . 35

3.2.1 Data Base Definition and Forms Design 36
3.2.2 Data Collection Techniques. 37
3.2.3 Keying of Data and Input Controls 42
3.2.4 Computer Edit Routines. 44
3.2.5 Manual Checks for Accuracy. 47

3.3 Summary of Weighting Factors 49

4.0 File Validation 52

4.1 Introduction 52
4.2 Validation Techniques. 52
4.3 File Validation Example. 53

5.0 Computer Simulation of Data Base Retrieval. 59

5.1 Precision and Recall 59
5.2 Query Formulation for Best Results 61
5.3 Soft-Match Techniques. 62
5.4 Example of Use of File Validation Summary Data 62
5.5 Precision and Probable Match 66

6.0 Conclusions and Recommendations 69

6.1 Conclusions. 69
6.2 Recommendations. 70

6.2.1 Evaluation of Current File Access Capability. . . 71
6.2.2 Evaluation of the Data Base 71

Appendix I - Precision Tables. 73

TABLE OF ILLUSTRATIONS

Page

Table 2.4-1 IBM Alpha Inquiry System Personal Name Encoding Algorithm. 15

Table 2.4-2 Western Airlines Surname Match Rating Algorithm. . . . 18

Table 2.4-3 Rules for Developing Phonetic Frequency Codes. 19

Table 2.4-4 Letter Distributions for Phonetic Frequency Code Name Lookup Technique 21

Table 2.4-5 Illustration of the Application of the NYSIIS Coding Techniques. 22

Table 2.4-6 NYSSIS Name Coding Rules 23

Form 3.1-1 Decay Factors of Standard Biographical Data Elements. 34

Form 3.2.1-1 Inherent Characteristics of Biographical Data Elements. 38

Form 3.2.1-2 Data Base Definition Weighting Factors for Biographical Data Elements 39

Form 3.2.2-1 Data Collection Procedure Weighting Factors for Biographical Data Elements 43

Form 3.2.3-1 Keying and Input Control Weighting Factors for Batch Input. 45

Form 3.2.3-2 Keying and Input Control Weighting Factors for Interactive Input. 46

Form 3.2.4-1 Computer Editing Weighting Factors 48

Form 3.2.5-1 Weighting Factors for Manual Accuracy Checks 50

Form 3.3-1 Cumulative Weighting Factors for Standard Biographical Data Elements 51

Table 4.2-1 Suggested Format for File Validation Record. 54

Table 4.3-1 Source Documents for Ten Sample Records from the Dummy Data Base. 55

Table 4.3-2 Computer Listings for Ten Sample Records from the Dummy Data Base. 56

Table 4.3-3 File Validation Record for a Sample Record from the Dummy Data Base 57

Table 4.3-4 Summary of Errors and Omissions Detected in Analysis of 10% of the File. 58

Table 5.3-1 Retrieval Ratios (Estimates of Proportion of File Retrieved) 63

Table 5.3-2 Statistical Classification of Surname Records as Compiled from Social Security Administration Accounts. 64

TABLE OF CONTENTS
(continued)

Page

Appendix II - A Probabilistic Formulation of the Non-Unique Access Problem 174

Appendix III - Soft-Match Techniques 181

Appendix IV - FORTRAN Source Code for the Probability Mode. . . . 186

ACCESSING INDIVIDUAL RECORDS FROM PERSONAL DATA FILES USING NON-UNIQUE IDENTIFIERS

G. B. Moore, J. L. Kuhns, J. L. Trefftzs, C. A. Montgomery
Operating Systems, Inc.

ABSTRACT

The Privacy Act of 1974 places restrictions on the Federal, state, and local agencies' use of the Social Security account number as an identifier. For some agencies, compliance will involve changes in implementation of retrieval algorithms. This report describes methodology applicable to these changes in the more general context of the problem of retrieving individual records from files using non-unique identifiers. State-of-the-art retrieval techniques are discussed, a method for assigning reliability weights to various personal data elements is presented, file validation techniques for the error and omission rates of data items are suggested, and a retrieval probability model -- designed to show likelihood of retrieval of a subject's record given a variety of populations, combinations of identifiers, and error/omission rates -- is described. A methodology is developed for forming confidence factors from the established error/omission rates for combinations of non-unique identifiers that are candidates for use as retrieval keys. Use of these confidence factors as indices into the precision tables produced by the probability model is described.

Key Words: Data retrieval; file validation; name lookup; non-unique identifiers; personal data files; Privacy Act, probability model; retrieval.

1.0 INTRODUCTION

1.1 The Legislation

The Privacy Act of 1974 opens with the words:

> "The Congress finds that-
>
> (1) the privacy of an individual is directly affected by the collection, maintenance, use, and dissemination of personal information by Federal agencies;
>
> (2) the increasing use of computers and sophisticated information technology, while essential to the efficient operations of the Government, has greatly magnified the harm to individual privacy that can occur from any collection, maintenance, use, or dissemination of personal information;
>
> (3) the opportunities for an individual to secure employment, insurance, and credit, for his right to due process, and other legal protections are endangered by the misuse of certain information systems;

> (4) the right to privacy is a personal and fundamental right protected by the Constitution of the United States; and
>
> (5) in order to protect the privacy of individuals identified in information systems maintained by Federal agencies, it is necessary and proper for the Congress to regulate the collection, maintenance, use, and dissemination of information by such agencies."[1]

There has been increasing public concern over the trend toward integration and centralization of data banks containing personal information. In the privacy legislation, Congress has explicitly defined the rights of citizens whose records are maintained by Federal agencies as well as the responsibilities of the agencies handling the data. In addition, the legislation also established the Privacy Protection Study Commission which is specifically authorized to:

> "...make a study of the data banks, automated data processing programs, and information systems of governmental, regional, and private organizations, in order to determine the standards and procedures in force for the protection of personal information... In the course of conducting the study...the Commission may research, examine, and analyze...the use of social security numbers, license plate numbers, universal identifiers, and other symbols to identify individuals in data banks and to gain access to, integrate, or centralize information systems and files."[2]

Specific prohibitions regarding use of universal identifiers will, for the most part, await the results of the Privacy Protection Study Commission study. The current emphasis is rather on restricting the data maintained to that directly relevant to the agency. However, since much of the concern over use of universal identifiers has been generated by the use of the Social Security account numbers and their potential for linking files, the Act has specifically provided against its use as a required item of information unless a Federal statute specifying such a disclosure is applicable.

> "It shall be unlawful for any Federal, State or local government agency to deny to any individual any right, benefit, or privilege provided by law because of such individual's refusal to disclose his social security account number."[3]

---

[1] Privacy Act of 1974, Public Law no. 93-579, Section 1.

[2] Loc. cit.

[3] Ibid., Section 7.

Thus the effect of the legislation is that individuals whose records are maintained in the files of Federal agencies must be allowed to review their own records on request, the retrieval must be possible without use of the Social Security account number, the data maintained in the files must be timely, and only data necessary in the performance of the agency's mission should be stored.

For agencies that have relied heavily on the Social Security number for purposes of retrieval, this requirement will necessitate some major readjustments. Further restrictions are imposed on the type of information that can be solicited from individuals who have requested copies of their record by the July 1, 1975 guidelines from the Office of Management and Budget, "Guidelines for Implementing Section 3 of the Privacy Act of 1974," which states that the items published in the Federal Register by each agency must include:

> "...the information necessary to identify the record. Where the system employs a specialized identification scheme, the individual should not be required to provide such a number or symbol as an absolute requirement, although the individual might be requested to supply it if he or she can reasonably be expected to know it. Instead, alternative combinations of personal characteristics may be used to identify individuals who may have lost, forgotten, or are unaware of their identification numbers or symbols. For example, the combination of name, date of birth, or place of birth, and father's first name may be sufficient to identify an individual without the use of a system identification number."[4]

Thus, although use of record identification numbers, driver's license numbers, etc., is not prohibited, careful attention must be given to the selection of a set of identifiers that will be considered reasonable in the particular environment.

1.2 Retrieval of Individual Records

The implications of these restrictions will differ from agency to agency. It is important, however, that those charged with responsibility for the selection and implementation of identification schemes be aware of the retrieval effectiveness of non-unique personal identifiers. The potential for retrieval by means of exclusive use of combinations of these keys should be understood, and the relevance of this potential for use with currently active data bases should be examined. Consideration must also be given to the question of the reliability of the data contained in these files since the success of the retrieval algorithm depends upon the accuracy of the data base itself. Errors may have been introduced into the data base (during data collection, data entry, data update, etc.), with the result that a request from an individual for a copy of his record might fail due to the presence of erroneous data in the record itself. In this respect, retrieval algorithms based on Social Security account number or other standard universal identifiers have advantages (it is easier to alert clerical staff to the necessity of checking one or two numbers for accuracy than it is to train them to be careful with all biographical data), adequate retrieval is certainly possible using non-unique identifiers such as name, birthdate, etc. In fact, such retrieval is sometimes possible even when the variables most closely identified with an individual are removed from the records. Statisticians involved in the collection and analysis of statistical data for research purposes have been concerned with the problem of protection of the individual whose data is maintained in the files. The need to collect such data in order to study causes of disease, poverty, violence, migration and other sociological phenomena is well understood, but the individuals involved have the right to expect that information given in confidence will be treated with integrity. Dr. Fellegi of the Dominion Bureau of Statistics in Canada has summarized the problem:

---

[4] Office of Management and Budget Circular A-108 and accompanying "Guidelines for Implementing Section 3 of the Privacy Act of 1974," Federal Register, 9 July 1975, Vol. 40, No. 32, pp. 28948-28978.

> "The public benefits indirectly from the legitimate uses of statistics by governments, businesses, non-profit organizations, academic users, etc.; yet, the public is also concerned about the increasing burden of providing the required statistics and about the real or imagined possibility of the misuse of the data provided by them. The explosive increase in the demands for more statistics can only be met, without impossible response burdens being put on the public, through a more effective exploitation of the data. This increases geometrically the magnitude of the problem of checking tabulations for disclosure."[5]

The seemingly contradictory requirements of protecting individual privacy while at the same time providing the citizen the benefits attendant on the introduction of automated record keeping systems create a burden for those responsible for maintaining the systems. They must restrict the data retained in the file to that clearly relevant to the application, but at the same time be able to identify correctly all individuals concerned on demand. Data must be available for research efforts, but information maintained to support these efforts must be restricted and accessing of individual records carefully monitored. The dilemma thus created is discussed further in the next section.

1.3 Statistical vs. Dossier Files

Until recently, it was generally believed that a basic distinction existed between statistical files (those with names, addresses, etc., removed from the individual record) and dossier files (those containing explicitly descriptive information about the individuals). Dossiers were considered dangerous from the point of view of exposure whereas statistical files were not. James Martin, in discussing this distinction, says:

---

[5] Fellegi, I. P., "On the Question of Statistical Confidentiality," Journal of the American Statistical Association, 67 (March 1972), p. 7-18

> "It is important to separate the concept of statistical files from the concept of dossiers about individuals. From the point of view of privacy protection, they are entirely different. Many of the uses to which the information in government and research data banks will be put are statistical. In statistical files the identification of the individual should be stripped off. Once that is done the files can be used more freely. It is still necessary to make sure that people cannot be identified by some other means, such as home location, a high salary, the fact that he has 13 children or cats, weighs 300 pounds, or has a combination of factors that makes him unique. It is sometimes possible to extract information about individuals from a statistical data bank by asking combinations of questions. Sometimes a large number of questions would be needed and controls can be devised to prevent such interrogation. The controls are more complex and subtle than on the type of data bank in which details about an individual are requested directly. In general, statistical data banks will not reveal the identities of the persons about whom data are recorded."[6]

Recent research indicates that the data contained in statistical files is not necessarily protected from unauthorized disclosure. Dr. Schlörer of the Department of Medical Statistics at the University of Ulm in Germany gives the following account of his investigation of retrieval from interactive data banks.

> "Until fairly recently, removing name, address, and -- perhaps -- date of birth from a record would secure anonymity in most cases. The advent of the computer has changed this situation. It was soon realized that the distinction between a dossier data bank (returning full identifying information to the user) and a statistical data bank (providing only numbers of persons) is largely delusive, if the user is allowed to communicate with the data bank by dialogue.
>
> A prerequisite for this intrusion by dialogue is preknowledge. The intruder must have some a priori information from the statistical data bank. The same information must be stored in Mr. X's record, so that preknowledge and the corresponding part of the record can be matched. The preknowledge must suffice to identify or, at least, nearly identify Mr. X. An identification experiment with authentic

---

[6] Martin, James, Security, Accuracy, and Privacy in Computer Systems, Prentice-Hall, New Jersey, 1973, p. 432.

> data will be described in this paper. It will be seen that relatively little preknowledge will do for identification purposes -- at least for the life 'attacked' in this experiment."[7]

Not only does Dr. Schlörer's work make it clear that records can be retrieved through the use of non-unique identifiers, it also gives added credence to the requirement that only relevant data be maintained in the files, since inclusion of extraneous data makes the record more vulnerable.

Further indication of the need for such restrictions comes from the research in record linkage carried out by Drs. Fellegi and Sunter. They have developed

> "...a mathematical model...to provide a theoretical framework for computer-oriented solution to the problem of recognizing those records in two files which represent identical persons, objects or events (said to be matched) ...A comparison is to be made between the recorded characteristics and values in two records (one from each file) and a decision made as to whether or not the members of the comparison-pair represent the same person or event, or whether there is insufficient evidence to justify either of these decisions at stipulated levels of error...A theorem describing the construction and properties of the optimal linkage rule and two corollaries to the theorem which make it a practical working tool are given."[8]

The existence of such techniques for linking individual records on the basis of non-unique identifiers again points up the feasibility of implementing retrieval algorithms based on those variables. An awareness of the availability of such techniques further clarifies the nature of the threats that currently exist to individual privacy. It also emphasizes the importance of the provisions for relevancy of data in the Privacy Act since obviously the more data available, the easier such linkages become. Similarly, the necessity for building audit checks into the system to provide a means for determining what information was accessed by whom becomes clear in the light of currently available methodologies.

1.4 Privacy vs. Security

Privacy has been perceptively defined by Westin as the right "...to determine what information about ourselves we will share with others,"[9] where the 'others' are persons or agencies who are presumably authorized to hold and use such information. Thus the legislative concept of privacy in effect concerns the protection of information about individuals from misuse by duly constituted authorities, or persons or agencies to which such authority has been delegated (e.g., for statistical studies).

---

[7] Schlörer, J., "Identification and Retrieval of Personal Records from a Statistical Data Bank," Methods of Information In Medicine, 14 (1, 1975).

[8] Fellegi, I. P., and A. B. Sunter, "A Theory for Record Linkage," Journal of the American Statistical Association, (Dec. 1969), pp. 1183-1210.

[9] Westin, Alan, as quoted by I. P. Fellegi, "On the Question of Statistical Confidentiality," op.cit., p. 18.

Security, on the other hand, is concerned with protection of this information from unauthorized 'others.' Such unauthorized access may be deliberate, or inadvertent: e.g., a person seeking information from his own record may be given the dossier of another individual by mistake.

Protection of information in a computerized file from unauthorized access is partly a technical, system-internal, matter, and partly a procedural, system-external matter. System-internal protections include obvious items such as use of passwords and other identifying techniques to guard against deliberate access attempts by unauthorized persons. More fundamentally, such protection requires a high degree of precision in the retrieval algorithm to prevent inadvertent delivery of the wrong information record to a particular requestor (as well as erroneous updates and purges).

Procedural protections involve basic items such as locks and security guards, as well as requests for proof of identity from persons seeking information about their own records, further verification based on questions about the content of the record before it is delivered to the requesting individuals, and so forth. In the case of a computerized information bank where requesting individuals are allowed access via a computer terminal, a combination of technical and procedural safeguards is required to insure protection of the component information records.

In a recent article on computer privacy and computer security, Willis Ware states:

> "In the context of computer-based systems, the matter of access control is an essential part of a larger issue referred to as computer security which can be defined as:
>
> > The protection of the equipment, facilities and data of an information system against deliberate or accidental damage, and against denial of use by legitimate users, together with the assurance that information will be delivered by the system only to individuals authorized to receive it."[10]

The phrase "legitimate users" of an information system implies in the privacy context that legitimate uses of personal information can be defined, and that "individuals authorized" to access such information will respect the right to privacy of those persons who provided it. This is the core of the privacy issue, which essentially involves the loss of control over personal information incurred by an individual in providing such data to some agency or data bank, and the profound effect that relinquishing control of this information can have on the life of the particular individual. It is the concern of those involved in the effort to protect the privacy of the individual to identify measures, legal, procedural, and technological, by which potentially detrimental effects of this loss of control can be nullified.

---

[10] Ware, Willis H., "Computer Privacy and Computer Security," Bulletin of the American Society for Information Science, Vol. 1, No. 3, Oct. 1974.

1.5 Summary

In Section 1 some of the difficulties faced by those responsible for the management of personal data bases in Federal agencies are outlined. Problems inherent in the provision of continued service with no attendant reduction in efficiency due to the increased emphasis on the protection of individual privacy have been stressed. The remainder of this report is concerned with the presentation of methodologies potentially useful in the minimization of these difficulties.

The application of current retrieval technologies to the problem of retrieval through use of non-unique personal identifiers is the basic methodology under consideration. A discussion of relevant retrieval algorithms is given in Section 2. Section 3 suggests an approach to determination of the accuracy and reliability of the various identifiers contained in current files. Forms that might be used in carrying out an analysis procedure are introduced, and a mechanism for arriving at weighting scores for the individual identifiers is described. Section 4 demonstrates a similar mechanism for analyzing error and omission rates for each of the personal identifiers contained in the file.

Section 5 describes the probability model developed to assist in determination of the best combination of personal identifiers available for use as retrieval keys given the idiosyncratic error and omission rates of those variables within a particular file. An example of the use of the output from the probability model in making this determination is given. Conclusions are formulated in Section 6 which also contains a summary of procedures recommended for those involved in the analysis of personal data files.

Four Appendices are included:

Appendix I: Precision Tables
Appendix II: A Probabilistic Formulation of the Non-Unique Access Problem
Appendix III: Soft Match Techniques
Appendix IV: FORTRAN Source Code for the Probability Model.

2.0 STATE-OF-THE-ART RETRIEVAL TECHNIQUES

2.1 INTRODUCTION

The personal data files affected by the provisions of the Privacy Act represent a wide spectrum of applications with very different requirements. By definition the files all contain dossier-type data, but they vary in terms of the non-biographical variables maintained, the ultimate use of the data, the frequency of file access, the retrieval mechanisms employed, the degree to which accuracy has been stressed, etc. These distinctions reflect the differing operational requirements of the agencies involved. Clearly, the provisions of the Privacy Act giving the individual access to his own records will cause re-evaluation of file management requirements and, in many cases, some system changes will be necessary in order to make this data available.

A clear-cut requirement for some system changes exists when the retrieval keys do not include any of those combinations of identifiers that can be readily recalled by the individual. Similarly, systems that have placed a high degree of reliance on the use of the Social Security account number in the retrieval algorithms will require extension of these system modifications to insure effective retrieval. But even if these changes result in the availability of a standard set of biographical variables as potential retrieval keys, the level of accuracy maintained for those variables may fall short of that required of retrieval keys. It is entirely possible that a higher degree of accuracy exists for the variables that have historically been used as retrieval keys than for those now promoted as candidates for that task. Such a disparity in levels of accuracy could have resulted from awareness on the part of the clerical staff of the need to verify the retrieval keys, or from the programming of special checks on those fields, or the like. Further, this skewing of levels of accuracy may well have gone unnoticed since the retrieval algorithm has operated sufficiently well to fulfill the operational requirements of the agency.

To illustrate this point, consider the case where an agency uses a specific data item, for example AGE, to trigger notification of the individual concerning a change in status. In implementing and maintaining such a system, the emphasis on manual and system checks of the variables BIRTHDATE and AGE would be greater than that on NAME and ADDRESS. Indeed, since postmen have become expert at detecting near matches on name and street numbers, the degree of accuracy required to complete the correspondence is indeed less than that required for retrieval of the record on the basis of current age.

One of the authors, Mr. Trefftzs, has experienced this phenomenon. His is an unusual name that causes a great deal of confusion. As a result it is seldom spelled correctly. Mr. Trefftzs has, in recent weeks, received communications from a variety of sources (governmental, magazines, professional societies, etc.) in which his name was consistently misspelled in a number of inconsistent ways:

TREVITZ TRAFFT TRAFFTZS TRREFTZA TREFTZ etc.

These misspellings prevented neither the retrieval of his record nor the delivery of the message. Yet, if the systems involved were capable of handling retrieval on the basis of biographical data including name, and he were to request a copy of his record from each, it is likely that some of the systems would fail to locate the record, given the nature of the error that exists in the name itself. This emphasizes the importance of the change of system perspective attendant on introduction of new accessing requirements. Problems such as gross misspellings of names may block retrieval of certain records until the level of accuracy of the NAME variable has been brought within reasonable bounds.

Existing file organizations and retrieval mechanisms may thus be entirely responsive to the needs of a particular agency while being inadequate in terms of handling the new requirements. The most difficult aspect of this problem arises from the fact that the required retrieval capabilities may not have been built into the original system. In some cases, extension of current system capabilities will be sufficient to handle the new requirements; in others, new sets of capabilities will be needed; still other situations will arise where restructuring of the files is required in order to make this data available in the required way.

Decisions regarding the system changes required will be based on several factors: demand (i.e., the number of requests for this information), capabilities of the current system (some will be set up in such a way that required changes will be minimal), characteristics of the data base (number of non-unique identifiers, available accuracy rates, typical problems with a given variable), etc. If it is evident that the complete redesign of the system is required in order to meet the new regulations, the agency will want to examine data base structures and retrieval techniques that might assist in optimizing the new system. The remainder of this section is devoted to a description of the current state-of-the-art for retrieval techniques as they apply to the problem of accessing data using non-unique identifiers. Since there is no single answer to the implementation of these requirements, this section is presented in the form of a survey in the hope that the availability of this information will clarify the options available to those confronted with the problem.

2.2 VARYING SELECTIVE VALUES OF NON-UNIQUE IDENTIFIERS

Some of the identifiers available in personal data files are highly individualized. The variable NAME is the most obvious example of this. Other data items for example, SEX, are binary, and therefore not particularly useful in identifying an individual record. Yet collection of even the most straightforward items can be difficult since the items requested are sometimes subject to vastly different interpretations.

There is a natural tendency to assume that basic biographical data is easily obtainable, and that standard definitions can be taken for granted. However, in actual data collection situations, surprising misunderstandings can and do occur and, most importantly, are often not detected until much later. The result is that variables considered to be reliable in the file are many times, in reality, only marginally so.

An interesting discussion of the problems inherent in obtaining biographical data is given in the Uniform Hospital Discharge Data Demonstration Summary Report:

> "DATE OF BIRTH: Date of birth is usually an easy item to collect on admission. However, there are instances of elderly persons not remembering their exact date of birth, and others who provide incorrect dates for personal reasons.
>
> SEX: The incidence of encountering a patient with true indeterminate sex status is very low. Most hospitals feel the indeterminate category is unnecessary, but hospitals who perform trans-sexual operations find it a useful item.
>
> MARITAL STATUS: While this item is usually easy to obtain, there are some inherent difficulties in assuring validity, and these include:
>
>> Persons living in common-law relationships who do not wish to be classified under any of the alternatives.
>>
>> Divorced persons who consider themselves single and state so at admission.
>>
>> Single females admitted for childbirth do not generally want to be identified as single and state they are married.
>>
>> The original intent of this item was to attempt to secure some indication of the living arrangement of a patient, but it proves to be of marginal utility for this purpose. This is a complex variable to establish if the many socio-economic factors which enter into the underlying health care implications associated with living arrangements are considered."[11]

---

[11] Hodgson, D. A., L. E. Kucken, and J. M. Ensign, The Uniform Hospital Discharge Data Demonstration, Summary Report, Health Service Foundation, Chicago, 1973, pp. 30-31.

2.3 IMPORTANCE OF THE NAME VARIABLE IN RETRIEVAL THROUGH USE OF NON-UNIQUE IDENTIFIERS

Use, as well as inherent variability, of identifiers will vary greatly from one application to the next. The variability within groups of identifiers is probably most extreme for the NAME variable. The following example of such variability comes from a bibliographic data base application. The report states that:

> "...many of the individual data elements exhibit great variety (i.e., lists of their contents are extensive), and show relatively disparate distributions. This behavior is encountered in different degrees in regard to items such as words in the titles of monograph or periodical articles, assigned subject headings, authors' names and citations. ...In general, the distributions are approximately hyperbolic, so that a small portion of items may account for a substantial proportion of occurrences, while the majority of items occur only infrequently. ...Of all the data elements, personal author names exhibit a distribution which is at its most extreme in one direction. As is shown ...the most frequent (full name) author name in a file of 50,000 names occurred only sixteen times, while over 35,000 of the names, or over 70 percent of the file, occurred once only."[12]

The same tables show that, with respect to surnames only 20,000 or 40% of the file, occurred only once. These figures emphasize the high selective value of NAME as an identifier and thus provide the rationale for the following consideration of name lookup techniques.

2.4 SELECTED STATE-OF-THE-ART NAME LOOKUP TECHNIQUES

A majority of the name lookup algorithms currently in use are based on a phonetic scheme aimed at minimizing the common transcription problems in the recording of names. The IBM Alpha Search Inquiry Program description states that:

> "Using the phonetic approach for personal name increases the user's ability to access the alpha search record even though the exact spelling of a name may not be known.

---

[12] D. W. Fokker and M. F. Lynch, "Application of the Variety-Generator Approach to Searches of Personal Names in Bibliographic Data Bases-Part I. Microstructure of Personal Authors' Names," Journal of Library Automation, 7 (2 June 1974), pp. 105-117.

> Thus the effects of transcription errors, partially illegible signatures on correspondence, and sound-alike names can be minimized."[13]

In analyzing name lookup techniques, it is useful to refer to the information systems concepts of precision and recall. Recall represents the number of relevant items contained in a given file which are retrieved in response to a particular query, while precision refers to the number of file items retrieved which are actually relevant to the query or search prescription.

In general, the objective of system improvement is to increase the degree of precision or reliability, without substantially increasing the amount of irrelevant material recalled or selected from the file.

In a document retrieval system, however, it is rare (in practice, at least) that there is a single file item which is capable of satisfying a given query. Generally speaking, several documents will be relevant to a query, where the degree of relevance is defined concretely in terms of the concepts used in the query and abstractly in terms of actual utility to the requestor.

Conversely, in a name retrieval system, the usual case is that there is a unique file item which satisfies a query, although the existence of other items listed under alternate spellings of the name must also be assumed by the search strategy. This is a critical issue, which does not appear to be effectively handled by many of the name search techniques investigated in the course of this study.

In evaluating name lookup techniques for which complete documentation is available, there are two essential criteria for determining relative effectiveness. Both of these may seem obvious, but since several of the lookup techniques described here fail on one or both counts, the topic merits discussion.

The first basic question is simply: does the strategy allocate similar names to the same logical group -- and more particularly, are alternate versions of the same name (Heinz, Heintze) in the same logical group? The answer is often in the negative, and this is the primary means of effecting an improvement in reliability of a given method, as we shall consider below.

The second basic evaluation criterion is the converse of the first: does the strategy allocate dissimilar names to different logical groups? Although the possibility of violating this principle appears unlikely, many name lookup techniques do in fact fail to satisfy this criterion, including Soundex, the first widely used phonologically based system for name encoding. Most of the systems currently in use have been influenced by the rules established 50 years ago by Margaret Odell and Robert Russell for the Soundex system.

---

[13] Alpha Search Inquiry System General Information Manual, GH20-1188-3, IBM Corporation, White Plains, N.Y., April 1974, p. 7.

Knuth summarizes the original Soundex rules as follows:

"1. Retain the first letter of the name, and drop all occurrences of a, e, h, i, o, u, w, y in other positions.

2. Assign the following numbers to the remaining letters after the first:

   b, f, p, v → 1
   c, g, j, k, q, s, x, z → 2
   d, t → 3
   l → 4
   m, n → 5
   r → 6

3. If two or more letters with the same code were adjacent in the original name (before Step 1), omit all but the first.

4. Convert to the form "letter, digit, digit, digit" by adding trailing zeros (if there are less than three digits), or by dropping rightmost digits (if there are more than three)."[14]

This algorithm fails to meet the evaluation criteria described above, as Knuth has pointed out:

> "...the names Euler, Gauss, Hilbert, Knuth, Lloyd, and Lukasiewicz have the respective codes E460, G200, H416, K530, L300, L222. Of course this system will bring together names that are somewhat different, as well as names that are similar; the same six codes would be obtained for Ellery, Ghosh, Heilbronn, Kant, Ladd, and Lissajous. And on the other hand a few related names like Rogers and Rodgers, or Sinclair and St. Clair, or Tchebysheff and Chebyshev, remain separate. But by and large the Soundex code greatly increases the chance of finding a name in one of its disguises."[15]

The type of refinement that has taken place in development of name look-up techniques is obvious from comparing the results of applying the rules used in the IBM Alpha Search Inquiry System (Table 2.4-1) with those of Soundex, using two of the examples mentioned above.

---

[14] Knuth, Donald E., The Art of Computer Programming, Vol. III: Sorting and Searching , Addison-Wesley, Reading, Mass., 1973, p. 392.

[15] Ibid.

Table 2.4-1 IBM Alpha Inquiry System Personal Name Encoding Algorithm [16]

This routine builds a phonetic key from the last name passed by the calling program. The phonetic key is 7 packed bytes (14 digits). The routine uses two tables, a first character table and a basic table. The first character table (see below) is used to recognize first letters or letter combinations and give them significance. For example, A in Armour has a value of 1 to be put in the key. A in Farmer has no value and is ignored. If the first letter or letter combination does not appear in the first character table, a zero is inserted in the first key position and the basic table is used to code the first character into the record key position.

The basic table (see below) is used to recognize letters or letter combinations that are phonetically equivalent. Vowels and the letters H, W, Y are ignored. Letters with the same phonetic value which are adjacent to each other are treated as single letters. For example, in CK and TT the second character is dropped. However: if a third character has the same value as the previous two, it is retained. The remainder of any unfilled key is filled with zeroes.

The first character table used in this routine is:

Letter(s)	Value
A	1
E	1
GF	08
GM	03
GN	02
H	2
I	1
J	3
KN	02
O	1
PF	08
PN	02
PS	00
U	1
W (except WR)	4
WR	04
Y	5

The basic table used in this routine is:

Code	Letter(s)
0	Z,S,CI,CY,CE,TS,TZ
1	D,T
2	N
3	M
4	R
5	L
6	J,SH,SCH,CH
7	C,G,K,Q,X,DG
8	F,V,PH
9	B,P

---

[16] Alpha Search Inquiry System General Information Manual, IBM Corporation, White Plains, New York, GH20-1188-3, 1974, pp.41-42.

Table 2.4-1 IBM Alpha Inquiry System Personal Name Encoding Algorithm (Continued)[16]

EXCEPTIONS TO BASIC TABLE PROCESSING*

In certain situations where letters or groups of letters have multiple sounds, a second and third pass through name encoding is made. Example: CH Has a hard sound in Nichols, a soft sound in Chavez.

	1st Pass (Coded in Record)	2nd Pass	3rd Pass
CZ	70	6	0
CH	6	70	0
CK	7	7	6
C	7	7	6
K	7	7	6
DS	0	10	10
DZ	0	10	10
TS	0	10	10
TZ	0	10	10

Following are some examples of the results of coding personal names:

NAME H A R P E R
KEY 2 4 9 4 = 24940000000000

NAME C O L L I E R
KEY 07 5 4 = 07540000000000

NAME S C H U L T Z
KEY 06 5 0 = 06500000000000

NAME L I V I N G S T O N
KEY 05 8 2 7 0 1 2 = 05827012000000

*Occurrences of the letter combinations in this table cause generation of alternate phonetic keys. When this happens, the lookup algorithm will use both phonetic keys in the search for potential matches. This procedure alleviates many of the problems attributable to phonetic misspellings (e.g., Ohrbock for Ohrbach). A cross-reference table is used to connected sound-alike names (e.g., Lyle and Lisle) which are not handled by this algorithm.

SEARCH CRITERION	SURNAME	ASSIGNED VALUES	
		SOUNDEX	ALPHA
Assign similar names to the same logical group	RODGERS	R326	04740000000000
	ROGERS	R262	04740000000000
Assign dissimilar names to different logical group	KANT	K530	02100000000000
	KNUTH	K530	07210000000000

In both these cases, the Alpha Search Inquiry System has solved the problem. There are, however, an equally good set of examples for problems that still remain. Thus, although this algorithm represents progress over that used in the Soundex system, it is far from perfect.

Sometimes quite straightforward algorithms are capable of handling this problem successfully for files of limited extent. The match rating approach used by Western Airlines (see summary in Table 2.4-2) would correctly identify RODGERS and ROGERS as a potential match. In the Western system, the matching algorithm would compare RDGERS with RGRS. The number of unmatched characters would be 1, the resulting similarity rating would be 5, and, hence, this record would be flagged for retrieval.

On the other hand, this algorithm would also consider KANT to be a likely match for KNUTH, which points up the limitations of the system. It is used successfully in this application because only a small file is searched at a time. Ms. Gail Hogan, Manager of Passenger Service Systems and Programming for Western, states:

> "Since our reservations system is flight/date oriented (i.e., we limit the scope of our search by date and/or flight number), we have little difficulty in obtaining an exact name match. However, if there are passengers on a flight with the same surname, we display a list of these to the agent. If an exact match is not made, we continue processing...in order to compile a similar surname list."[17]

Another example of failure to meet the criterion of assigning dissimilar names to different buckets can be given using the Standardized Phonetic Frequency Code algorithm (see Table 2.4-3) as described in the NYSIIS evaluation of name search techniques.[18] In this system, an attempt was

---

[17] Personal Communication from Gail Hogan, Manager Passenger Service Systems and Programming for Western Airlines, September 15, 1975.

[18] Taft, Robert L., Name Search Technique, New York State Identification and Intelligence System, August 1970.

Table 2.4-2 Western Airlines Surname Match Rating Algorithm[19]

The input surname and each of the PNI [Personal Numeric Identifiers] - item surnames are encoded by this program prior to comparison. Encoding is as follows:

(1) Deletion of all vowels unless the vowel is the first character of the surname.

(2) Elimination of all double consonants by deleting the second contiguous usage of any consonant.

(3) Reducing all encoded names to a maximum of six characters. This is done by retaining the first three and last three encoded characters.

The lengths of each pair of encoded names (input and PNI item) are examined.* If they differ in length by more than two, no similarity comparison is performed. A minimum acceptable similarity rating is established for each pair of encoded names as follows:

Sum of lengths is 4 or less; rating of 5
Sum of lengths is 7 or less; rating of 4
Sum of lengths is 11 or less; rating of 3
Sum of lengths is 12; rating of 2

The minimum acceptable rating establishes which PNI surnames are not "similar enough" to be considered by the agent.

Comparison of encoded names is then performed. This comparison is from left to right, character by character. Matching pairs of characters are deleted. This comparison continues until either encoded name has no remaining characters. All unmatched characters in both encoded names are packed to the right and comparison proceeds from right to left. On completing these comparisons the number of unmatched characters in the longer name is subtracted from a value of six with the result being the similarity rating for that PNI item.

Each PNI item that has a similarity rating equal to or greater than the minimum rating established for this time will be added to the Retrieval Control Record [RCR]. This program will add to the RCR until the record is filled. At that point additional entries will replace other items in the RCR which have lower similarity ratings. If no item with a lower rating is found, the PNI item will be ignored.

*That is, the length of the encoded input name is compared with that of the encoded name in the file.

---

[19] Gail Hogan, op.cit.

Table 2.4-3 Rules for Developing Phonetic Frequency Codes

		NAME FIELD	CODE
		J.KUHNS G.ALTSCHULER	
1.	In the name field, convert DK to K, DT to T, SC to S, KN to N and MN to N.	J.KUHNS G.ALTSHULER	
2.	In the name field, replace multiple letters with a single letter.	J.KUHNS G.ALTSHULER	
3.	Remove vowels, W, H, and Y but keep the first letter in the name field.	J.KNS G.ALTSLR	
4.	The first digit of the code is obtained using PF1*and the first letter of the name field. Remove this letter after coding.	J.NS G.LTSLR	1 3
5.	Using the last letters of the name, use Table PF3*to obtain the second digit of the code. Use as many letters as possible and remove after coding.	J.N G.LTSL	16 35
6.	The third digit is found using Table PF2*and the first character of the first name. Remove after coding.	N LTSL	167 357
7.	The fourth digit is found using Table PF2*and the first character of the name field. If no letters remain use zero. After coding remove the letter.	 TSL	1676 3579
8.	The fifth digit is found in the same manner as the fourth using the remaining characters of the name field, if any.	 SL	16760 35797

*Defined in Table 2.4-4

made to achieve a uniform distribution of codes across a set of logical groups. But in the actual implementation, some phonetically dissimilar symbols were assigned the same Phonetic Frequency code value as shown in Table 2.4-4.

Applying the rules shown in Table 2.4-3, the names J. Kuhns and G. Altshuler are reduced to the codes 16760 and 35797, respectively. However, as shown by the groups of letters associated with the same codes in Table 2.4-4, 'T. Vines' would also reduce to 16760, and 'J. Butler' to 35797. Moreover, although these quite dissimilar names would be represented by the same code, 'Kuntz' -- a variant of 'Kuhns' -- would not.

Returning to the issue of coding similar names such that they are allocated to the same logical group, it is clear that this principle is especially critical for alternate spellings of the same name. This logic is based on the assumption that non-detectable* errors in recording the name of 'Kuhns' may write 'Kuntz' instead. Also, persons changing their name to a more anglicized orthography and having file records listed under both spellings should be more easily traceable if additional phonological-to-orthographic coding rules were introduced.

The NYSIIS coding technique introduces a few such rules -- e.g., KN → N, Z → S, as well as a number of ad hoc rules to achieve greater reliability for the New York State Criminal justice name inventory. However, many of these rules are especially aimed at a large population of Spanish surnames and may at best not be useful for files with small populations of Spanish surname records. At worst, they may introduce undesirable ambiguities, as in the case of 'Risque' in Table 2.4-5, which illustrates the application of the NYSIIS coding rules given in Table 2.4-6 to a list composed mainly of OSI employees.

Again, 'Kuhns' and 'Kuntz' would result in different codes, whereas the dissimilar Chinese surnames Li, Lu, Low, Liu, Lao would all result in the same code, 'L'.

There is obviously still room for improvement, via the introduction of phonological-to-orthographic rules based on linguistic -- rather than ad hoc -- considerations. Moreover, such rules should be developed with a view to accommodating a variety of ethnic surnames used in the United States without biasing the coding techniques such that rules derived to handle one type of surname cause ambiguous coding of other surname types.

---

*This is based on the assumption that most typographical errors can be detected by comparing new names to the existing name list for a file and generating a computer listing of unique occurrences for manual inspection. This of course assumes an investment of computer and human resources for error detection and correction.

Table 2.4-4 Letter Distributions for Phonetic Frequency Code Name Lookup Technique

PHONETIC FREQUENCY GROUP	CODE VALUE ASSIGNED									
	0	1	2	3	4	5	6	7	8	9
PF1	S	C	F	A	L	D	E	G		
	Z	K	P	B	O	H	M	J		
		Q	U		R	I	N	T		
		V	W				X			
PF2	S	C	F	A	O	D	M	G	U	E
	Z	K	P	B	R	H	N	J	V	L
		Q	X			I		T	W	
PF3	B	D	F	G	M	R	S	Z		
	C	T	L	J	N		Z	E		
	K		P	X				H		
	Q							I		
	V							O		
								U		
								W		
								Y		
								MN	STN	STR
								TR	SN	SR
								DRS	PRS	TN
										TD

Table 2.4-5 Illustration of the Application of the NYSIIS Coding Techniques*

	(1)	(3)	(5a)	(5b)	(5e)	(5f)	(6)	(7)	(9)	RESULT
WORTHY		W	WARTHY		WARTY					WARTY
KUHNS	CUHNS	C	CAHNS		CANS			CAN		CAN
REITZ		R	RAATZ	RAATS			RATS	RAT		RAT
TU		T	TA						T	T
OGATA		O	OGATA						OGAT	OGAT
WHELCHEL		W	WHALCHAL		WALCAL					WALCAL
MONTGOMERY		M	MANTGAMARY	MANTGANARY						MANTGANARY
REVENTLOW		R	RAFANTLAW			RAFANTLAA	RAFANTLA		RAFANTL	RAFANTL
RISQUE		R	RASQAA	RASGAA			RASGA		RASG	RASG
COSTALES		C	CASTALAS					CASTALA	CASTAL	CASTAL
TREFFTZS		T	TRAFFTZS	TRAFFTSS			TRAFTS	TRAFT		TRAFT

Table 2.4-6 NYSIIS Name Coding Rules

1. If the first letters of the name are
   'MAC' then change these letters to 'MCC'
   'KN' then change these letters to 'NN'
   'K' then change this letter to 'C'
   'PH' then change these letters to 'FF'
   'PF' then change these letters to 'FF'
   'SCH' then change these letters to 'SSS'

2. If the last letters of the name are
   'EE' then change these letters to 'Yb̸'
   'IE' then change these letters to 'Yb̸'
   'DT' or 'RT' or 'RD' or 'NT' or 'ND'
   then change these letters to 'Db̸'

3. The first character of the NYSIIS code is the first character of the name.

4. In the following rules, a scan is performed on the characters of the name. This is described in terms of a program loop. A pointer is used to point to the current position under consideration in the name. Step 4 is to set this pointer to point to the second character of the name.

5. For each successive position of the pointer, only one of the following statements can be executed.

   a. If blank go to rule 7.
      if the current position is a vowel (AEIOU)
      and if it is E followed by V then change to 'AF'
      otherwise change current position to 'A'.

   b. If the current position is the letter
      'Q' then change the letter to 'G'
      'Z' then change the letter to 'S'
      'M' then change the letter to 'N'

   c. If the current position is the letter 'K'
      and if the next letter is 'N' then replace
      the current position by 'N' otherwise replace
      the current position by 'C'

   d. If the current position points to the letter string
      'SCH' then replace the string with 'SSS'
      'PH' then replace the string with 'FF'

   e. If the current position is the letter 'H' and either
      the preceding or following letter is not a vowel (AEIOU)
      then replace the current position with the preceding letter.

Table 2.4-6 NYSIIS Name Coding Rules (Continued)

f. If the current position is the letter 'W' and the preceding letter is a vowel then replace the current position with the preceding position.

If none of these rules applies, then retain the current position letter value.

6. If the current position letter is equal to the last letter placed in the code then set the pointer to point to the next letter and go to Step 5.

   The next character of the NYSIIS code is the current position letter.

   Increment the pointer to point at the next letter.

   Go to Step 5.

7. If the last character of the NYSIIS code is the letter 'S' then remove it.

8. If the last two characters of the NYSIIS code are the letters 'AY' then replace them with the single character 'Y'.

9. If the last character of the NYSIIS code is the letter 'A' then remove this letter.

The NYSSIS coding rules are, in several instances, subject to different interpretations. As a result, implementation in software may differ slightly from one system to another. This illustration demonstrates the results of one such implementation.

An example of a pragmatically developed name lookup technique is that in use at the Medical Information Bureau. In this system, neither a Soundex nor a phonetic technique is used. Robert W. Keighley, Director of Systems and Planning at the Recording and Statistical Division of Sperry Rand, the servicing agent for the MIB, describes the technique as follows:

> "The MIB system is a completely automated on-line batch oriented name look-up system in contrast to a visual terminal inquiry response system. The processing of inquiries is not accomplished through use of an algorithm such as a Soundex or phonetic technique; instead, a name key is generated (group spelling number) through use of our name dictionaries. These dictionaries of surnames and given names reside on a disc file. The dictionary is used only to convert the incoming surname or given name to a unique group number. This group number represents names which are confused with one another. In addition, for given names, alternate groups are generated for situations such as nicknames and other confusions which do not necessarily sound like the given name in the inquiry.
>
> In contrast to other popular algorithms which attempt to accomplish the above, but do not fully, we have manually grouped our names according to our experience over the last eight years. This has come about by having to either search or update at least 160 million records over this span of time. Needless to say, our dictionaries are fairly complete at this point in time."[20]

This technique, used in conjunction with a highly specialized search strategy (discussed later in this section), has given excellent results. Mr. Keighley states that:

> "In order to give you some measure of our success, we handle all the situations above (compound names, nicknames, cross reference names such as aliases and maiden names, etc.) against a file of 14 million records, processing an average of 80,000 inquiries a day. It should, also, be pointed out that no manual intervention, either at our computer site or at our subscribers location is needed prior to conveying the results of our search to the end user, since we identify the individual with a high degree of accuracy."[21]

---

[20] Personal Communication from Robert Keighley, Director of Systems and Planning, Recording and Statistical Division of Sperry Rand, the servicing agent for the MIB, October 14, 1975.

[21] Ibid.

This example illustrates the feasibility of developing a highly successful retrieval system based on a name lookup technique. But the high degree of special treatment given to development of this system was a time-consuming and very expensive process. It is equally important to point out that successful systems based on modified Soundex rules are also available: for example, the retrieval system in use at the Department of Justice in the Immigration and Naturalization Service. In other words, although the MIB system represents a high level of attainment, there exist algorithms within the public domain which can be used very effectively.

A fundamental consideration associated with the adoption of any technique is of course the tradeoff between computer costs for implementing and executing complex algorithms, human resources for monitoring computer activities, and the cost of missing a record in a search, which may have social and legal implications. In view of the latter, it seems more reasonable to accept larger selectivity ratios based on name lookup, assuming that other non-unique attributes can be effectively used to filter lists of file items retrieved by surname lookup.

2.5 IMPLEMENTATION OF NAME LOOKUP TECHNIQUES

The high selectivity of value of name as a non-unique identifier has prompted the emphasis on name lookup techniques. But it is also true that this emphasis has repercussions regarding selection of retrieval techniques. Some of the more sophisticated techniques will not be easily available due to the nature of possible errors in the data. For example, it is doubtful that hashing techniques would be useful in the processing of surnames since a misspelling either at data entry or retrieval request time could result in a complete misdirection of the search.

Use of name lookup techniques facilitates grouping of closely related names. Thus, if SMITH and SMYTHE are encoded in such a way that a sort of the file on the encoded field places them close to each other, the likelihood of finding the alternate spelling during a search is greatly enhanced. For example, if the Soundex code appears to be adequate for handling the mix of names in the file, the approach taken could be as follows: apply the Soundex algorithm to the SURNAME field in all records, perform a sort on the encoded field, and write the records, in sort order, on a storage device. The records will be recorded in convenient logical groups variously called blocks or buckets. (Both terms imply the number of physical records involved in a single read or write operation.)

If the procedure described above were followed, it would be likely that SMITH and SMYTHE would be recorded in the same, or at least in adjacent, buckets. Indices would then be formed for use in locating buckets during a search. In a small file, there might be a single level of index based, for example, on the Soundex formulation of the input surname. The first two or three alphanumerics of the resulting code could point to a bucket that would be expected to contain the desired record. In a larger system, a multi-level index scheme might be used with the highest level pointing to a secondary index containing a breakdown of names starting with the letter 'S.' In all cases, the indicated bucket will be read in and a matching algorithm followed until either an exact match is found or the contents of the bucket have been exhausted. If no match is found, the search may continue through use of the preceding and subsequent buckets. In some cases, it may be reasonable to look in logical blocks that are two or three buckets away since in the case of very common names the instances may stretch over several buckets.

The system developed by the Immigration and Naturalization Service follows this type of approach with the Soundex codes producing the top level index. The resulting groupings are then further segregated on the basis of the first names. The third level of index, BIRTHDATE, gives flexibility to the retrieval system. The program is capable of handling requests for the records of individuals of exactly this age or for the records of those within bounds of plus or minus nine years. The user has the option of asking for a count of hits before results are printed in case an unmanageable number of records have been retrieved due to specification of a wide age range.[22a] Obviously this option is for use by inquirers who are not sure of the individual's exact age and not by the individual requesting his own record. However, the number of problems than can occur with the AGE/BIRTHDATE variable make it not unlikely that an individual could be mistaken about how his own age has been recorded. The handling of such an error could be simplified by the availability of a range field.

The search mechanism employed by the Immigration and Naturalization Service necessitates use of three levels of indices before the actual accessing of the record(s). The number of required index or file searches can be reduced through use of encoding techniques to form a compound pointer. A clever approach to using several attributes as a record-key index or compound pointer was developed for the Central California Red Cross Blood Center. The program, whose main purpose is to maintain and develop a donor pool of repeat donors, was in operation at Stanford University at the time of this study.

> "Donor records are...stored in contiguous blocks by blood type, 30/block, and the donor indexes that point to these records are stored 255/block. By this method of indexing, the indexes themselves can be sorted by blood type, date of last donation, etc. relatively easily so that one set of donors can be dealt with individually."[22b]

[22a] Personal Communication from Robert Robinson, Director of ADP Systems Branch, Immigration and Naturalization Service, Department of Justice, September 22, 1975.

[22b] Ludwig, H. R., "An Interactive Computer-Based Donor Management System," Computers in Biology and Medicine, 5 (1975), pp. 69-75.

The donor index contains the following attributes:

- a two-byte donor i.d. consisting of the first three letters of the last name and the birthdate. This information is reduced to two bytes through use of the following algorithm:

  $$\text{Key} = \sum_{i=1}^{3} (a_i*26^{3-i})+J-2400$$

  where a_i denotes the numerical value of each of the first three letters of the last name, e.g. A = 1,B = 2,...and J denotes the Julian value of the birthdate with the base being 1900.
- a two-byte pointer that specifies both the physical block and the logical record number. For the current file, this is a number between 1 and 29,940. The organization of the file is such that this number also implies blood type.
- a two-byte data field to indicate date of last donation. To reduce field size, this is a Julian date from some base year (in this case, 1972).
- two one-byte fields for coded information: number of entries on the eligibility list, total donations to date, indication of rare blood types, etc.

The system has been designed so that updating of some levels of information can be handled by reading in only the index level, and, when a donor record is required, it can be fetched directly.

A very sophisticated example of use of an encoding technique occurs in the MIB system. This system places a high premium on accuracy of retrieval. Dictionary lookups are performed to obtain group numbers both for surname and given name. Mr. Keighley has summarized the search algorithm as follows:

> "The process of looking up an individual using surname, given name, date of birth, place of birth (optional) and territory of residence (optional) is as follows:
>
> 1. Look up surname in surname dictionary to obtain its group number.
> 2. Look up given name in given name dictionary to obtain its group number and any alternatives (nicknames etc.).

3. Generate a key consisting of surname group number, given name group number, date of birth (coded), place of birth (optional-coded), and territory (optional-coded).

4. Using the generated key from Step 3, search the data base on a direct access basis (index sequential) for records applicable to the inquiry.

5. Screen each record in detail and mathematically determine what the degree of probability is that the record being screened pertains to the person being searched.

6. If any record being screened has a high enough probability, it is selected as a response to the inquiry."[23]

In this instance the encoding technique has not cut down on the number of required accesses, but has instead been used to generate a very accurate filter.

2.6 OTHER SEARCH STRATEGIES

Analysts considering potential system changes will do so in the context of the operational environment of their agencies. In some instances, no great emphasis will be placed on search strategy. Yet, in other situations, particularly when the file size is large, the selection of an efficient search strategy will be crucial to effective system performance. The number of requests to be processed, batch/interactive capabilities of the system, size of the file, etc., will be determining factors in these decisions. A general discussion of some available search strategies is given here as background for those who must concern themselves with this issue.

W.A. Burkhard and R.M. Keller have described an approach to the problem of "...searching the set of keys in a file to find a key which is closest to a given query key...Three file structures are presented together with their corresponding search algorithms, which are intended to reduce the number of comparisons required to achieve the desired result."[24] They describe their result as follows:

> "The study described here is concerned with the general problem of efficiently searching files in the following situations: (1) find a key in the file closest to a given query key, and (2) find all keys in the file

---

[23] Keighley, op.cit.

[24] W. A. Burkhard and R. M. Keller, "Some Approaches to Best-Match File Searching," Communications of the ACM, 16, (4 April 1973), p. 230.

> closest to a given query key. By 'closest' we mean according to some suitable measure of distance, specifically a 'metric' in the mathematical sense. These situations may arise in information retrieval applications in which members of a file are keyed to a number of numerically-represented attributes. Our search algorithms have the following common basis: let b be the 'best' key found at a certain point in the application of the algorithm. Based on certain 'cutoff criteria' which depend on b, subsets of the file remaining to be examined are eliminated without explicitly comparing the keys in these subsets with the query key. Our experimental results (on randomly-generated files) indicate that large segments of the file may be eliminated from consideration using these cutoff criteria."[25]

Abraham Bookstein has described a hybrid search technique which "...has the advantage of retaining the simplicity of search keys while also including some of the flexibility that Boolean expressions of key words have for uniquely defining an item. ...The only indexes that must be maintained are the hash tables; the other indexes...are replaced by the search algorithms."[26] Bookstein goes on to say:

> "A user would begin by entering into the system search key. If the system finds that the number of items that would be retrieved exceeds a preset threshold, it would output a message requesting that the user enter a set of key words taken from various fields in the records; ...The system first generates a subfile of records having the desired search key. If a hashing technique is used, constructing this subfile can be accomplished quickly and at a relatively little cost in space for tables. Once the smaller file is formed, a complete search of the full records can be made for the key words. Since the system operates in two phases, it is less sensitive to the number of records the search key retrieves as far as user considerations are concerned. Ease of use becomes the dominating objective in designing the search key ... In the hybrid system, we can think of the search key not as an access mechanism ... but rather as a file reduction mechanism. This system trades the cost of maintaining and storing large indexes for an increase in costs of computer processing; only relatively easily maintained hash tables for fixed length search keys need to be maintained."[27]

---

[25] Ibid., p. 236.

[26] Abraham Bookstein, "A Hybrid Access Method for Bibliographic Records," Journal of Library Automation, 7:97-104, (2 June 1974), p. 99.

[27] Ibid.

The use of Malcolm Harrison's hash technique to expedite the search is explained as follows:

> "A fixed number of bits, or signatures, (is) added to each field on which a search can take place; these additional bits are derived in a well-defined way from the original field. This subfield is a fixed-size representation of the full field in a form that can be used to very rapidly eliminate most records which would not pass the key word matching test. It is stored in the index to the file along with the address of the record. Though this preliminary test is not foolproof, it could considerably reduce the size of the subfile that requires a more costly complete search, thereby reducing the number of disc accesses."[28]

The author quotes a false drop rate of approximately ten percent for the bibliographic application being described. This is of course too high a rate for personal data file retrieval systems, yet a careful implementation of a similar combination of such techniques might be appropriately applied to the retrieval problem at hand.

Retrieval techniques described in this section were selected because of their relevance to the present discussion of use of non-unique identifiers in search algorithms. Those interested in pursuing a broader range of retrieval techniques might begin with Dennis Severance's generalized model for exploring the relationship between alternative search techniques[29] and Knuth's work on sorting and searching[30].

---

[28] Ibid.

[29] Dennis G. Severance, "Identifier Search Mechanisms: A Survey and Generalized Model," Computing Surveys, 6 (3 September 1974), pp. 175-194.

[30] Knuth, op.cit.

3.0 IDENTIFICATION OF WEIGHTING FACTORS

3.1 ACCURACY CONSTRAINTS IMPOSED BY THE PRIVACY ACT

One of the most important facets of the privacy legislation is the imposition of accuracy constraints on the agencies maintaining personal data files. The intent of these provisions is to protect the individual's privacy by insisting that data pertaining to him be correct, up-to-date, and relevant. The new law gives the individual the right to review his own record and to request that any relevant corrections be made. Disputed information maintained in the file must be so identified.

This emphasis on accuracy of personal data held in computer files is a function of data availability, not of an increase in error rates due to computerization of the data. Manual record-keeping systems have, in general, a higher error rate than computer systems. Susan Wooldridge, et al. have stated that an error rate of up to 10% is not unusual in paper-based files. "Indeed a frequently used justification for computerizing is to raise the accuracy of the files."[31]

Alan Westin has quoted Dr. Collen of Kaiser-Permanente as indicating that manual systems in the hospital setting are plagued by errors.

> "Errors have become so common in traditional medical records that physicians are accustomed to handling them, from finding a woman's lab test in her husband's file, to prescription errors in the hospital."[32]

Dr. Collen went on to say that, after installation of a computerized record-keeping system, they had found that:

> "...some 10 percent of our computer records contain 'errors.' We don't know what the error level in manual files is, but it certainly couldn't be less. Gradually we hope to get the computer file error level down to around three percent. To get to 0 percent would be prohibitively costly, and probably isn't possible, anyway."[33]

Before the emergence of computerized files, few organizations took time to study their own systems from the point of view of accuracy and possible means of improvement.[34] There was far less concern with the problem, to a great extent because clues other than the data itself were readily available. These clues took the form of the handwriting in which data was recorded, the color of the form used, the placement of the paper in the file, the age of the document, etc. In many cases, the person reviewing the data was conversant with the general structure of the file, and often even with the individual case. He thus brought his own background to the perusal of the material.

---

[31] Wooldridge, Susan, Colin Corder, and Claude Johnson, Security Standards for Data Processing, John Wiley and Sons, New York, 1973, p. 60.

[32] Westin, A., Databanks in a Free Society. Quadrangle Books, 1972, p. 210.

[33] Ibid.

[34] Ibid, p. 433.

Automation has had the effect of depersonalizing the data, reducing it artificially to a level of uniformity that masks the disparities in accuracy while at the same time making it readily available to a broad spectrum of users. The loss of visual clues such as ancient appearance of the document or illegibility of the data is particularly important from the point of view of timeliness of the data maintained in the file.

James Martin, in his book entitled Security, Accuracy, and Privacy in Computer Systems, has discussed this problem.

> "In addition to patent errors, large files also accumulate two other kinds of troublesome material--the decaying and the unevaluated. The former is characteristic of biographical details. Information gets less reliable the farther away it is from the source. References and testimonials, to say nothing of college degrees, lose their relevance as their subject grows older. The facts remain facts in the environment of their collection, but they may be inappropriate in the context of the use to which they are put."[35]

The decay factor inherent in much biographical data might be thought of as the length of time during which the data is more likely to be right than wrong. For some variables, such as birthplace, there is no decay. For others, such as current age, decay is at a constant rate. For still others, the amount of decay will depend upon the time at which the data was collected, e.g., the variable HIGHEST DEGREE ATTAINED will have a much greater potential for change if it refers to a student than if it refers to a professional person.

Form 3.1-1 identifies some of the standard biographical data items that are subject to this type of variability. The privacy legislation has emphasized the importance of the timeliness factor since clearly much of this data deteriorates over time. In the context of the current discussion, this factor is an important indicator in the assignment of weighting scores to be used in a final determination of selection of non-unique identifiers for retrieval algorithms.

---

[35] Martin, J., Security, Accuracy, and Privacy in Computer Systems, Prentice-Hall, New Jersey, 1973, p. 431.

Form 3.1-1 Decay Factors of Standard Biographical Date Elements

NON-UNIQUE IDENTIFIER		DECAY FACTOR Does this variable remain constant over time?	
		YES*	NO**
LAST NAME	MALE	✓	
	FEMALE		✓
First Name		✓	
Middle Name		✓	
Sex		✓	
Birthdate		✓	
Birthplace		✓	
Marital Status			✓
Address			✓
Occupation			✓
Education			✓
Physical Characteristics		✓	

*Changes are rare
**Changes are not unusual

3.2 DETERMINATION OF WEIGHTING FACTORS FOR EXISTING FILES

In the establishment of a computer file, errors can be introduced at any stage of data handling. They sometimes result from the individual's misunderstanding of a question, sometimes it is a transcription error on the part of the clerk filling out a keypunch form, sometimes it is a problem in the file management system that creates or updates the record. A determination of the likelihood of such errors having occurred requires evaluation of procedures followed during the key stages in data handling:

1. data base definition and forms design
2. data collection techniques
3. keying of data and input controls
4. computer edit routines
5. manual accuracy check of record as printed by the computer

Thus, an acquaintance with the evolutionary process through which the file came into existence is essential to an evaluation of the degree of accuracy of the data it contains. A comprehensive examination of this evolutionary process should answer such questions as: is the data contained in the files consistent? (i.e., did all of the individuals who filled out the forms understand the same thing by the questions? Was the question asked the same way each time or has the wording changed significantly?) and: which data items in the file are likely to be most reliable?

The following sections describe each stage of data collection in terms of appropriate error detection procedures. Acknowledgement of the possible existence of such errors is important, but, since data base deficiencies will differ greatly from system to system, it would be helpful to be able to characterize these weaknesses in more specific terms. One approach would be to assign weighting scores to the various biographical identifiers in the file in such a way that the weights serve as indicators of the adequacy of the collection procedures for each variable. In the following discussions, a vehicle for making such a determination has been suggested. Included with the description of each stage of data collection is a table in which YES/NO questions are postulated for each of the problem areas identified. These questions are to be asked for each of the biographical identifiers used in the hypothetical file under consideration. A summary chart is used to record the total number of NO checks for each of the data elements. These totals then become the identified weighting scores for the various data items and thus produce a gross picture of the validity of the biographical identifiers in the file.

The weights thus arrived at for each of the standard biographical data elements will be used in a final determination of identifiers for the retrieval process (see Section 4).

3.2.1 Data Base Definition and Forms Design. The determination of an appropriate set of data items to be included in a system is an important part of the design. If the items requested are inappropriate or the wording of the question obscure, difficulties arise which are not obvious from looking at the data. The importance--and the difficulty--of obtaining such uniformity of data is discussed in the Uniform Hospital Discharge Data Demonstration report. This project was directed toward improvement in health care delivery through collection and evaluation of discharge data from a large number of hospital facilities throughout the country. A subcommittee of the U.S. National Committee on Vital and Health Statistics handled selection of the basic data set. Collection of the data was carried out systematically, with personnel at the hospital sites being trained with respect to abstracting of data items from the charts, and quality control checks implemented throughout the data collection procedure.

> "In all instances, the system manuals used in the test hospitals were modified to conform the UHDDD manual...Coding examples were provided to elaborate on the individual items and to cover anticipated contingencies such as when the date of birth, ZIP code, or married status was unknown, when patients from foreign countries who do not have ZIP codes were treated, etc. Examples to clarify the concepts of the principal diagnosis and the principal procedure were also given."[36]

Even so, many data collection problems have occurred. For example, the categories established for MARITAL STATUS are:

> "Now Married refers to persons who state they are married and not separated. Divorced or Separated is used for persons who state they no longer live together, whether or not legal action has been taken. Never Married (Single) refers to persons who state they have never married or whose only marriage has been annulled. Widowed includes widows and widowers."[37]

Yet, as noted in Section 2.2, there were difficulties in collecting even this very well-defined data in the following instances:

> "Persons living in common-law relationships who do not wish to be classified under any of the alternatives. Divorced persons who consider themselves single and state so at admission. Single females admitted for childbirth do not generally want to be identified as single and state they are married."[38]

---

[36] Hodgson, D. A., L. E. Kucken, and J. M. Ensign, The Uniform Hospital Discharge Data Demonstration, Summary Report, Health Services Foundation, Chicago, Illinois, 1973, p. 24.

[37] Ibid, p. 12.

[38] Ibid, p. 31.

They conclude:

> "The original intent of this item was to attempt to secure some indication of the living arrangement of a patient, but it proves to be of marginal utility for this purpose. This is a complex variable to establish if the many socio-economic factors which enter into the underlying health care implications associated with living arrangements are considered."[39]

Form 3.2.1-1 contains an analysis of the inherent characteristics of a sample set of biographical data elements. The results indicated in the form reflect good experience, but it is assumed that counter instances could be found. The analyst should compare his own evaluation of similar elements contained on the forms used in his data collection procedures with these results.

The data requested in Form 3.2.1-2, unlike that shown in the earlier forms, is dependent upon the actual collection procedures involved and is, therefore, left empty. The analyst is encouraged to produce similar tables, using the identifiers relevant to his file, and then, using a selection of random records from the data base in conjunction with the forms on which the original data collection took place, to answer the questions in the table. The answers will supply the required weights for this stage of the data collection procedure.

3.2.2 Data Collection Techniques. Perhaps the best possible guideline for successful data collection would be: capture the data as close to the source as possible. This might equally well be stated as: the fewer manual transcriptions of the data, the cleaner the data. The reasoning behind this is that the individual is normally the person best qualified to monitor collection of information about himself. He would not be as likely, for example, to transpose numbers in his street address as would a clerk whose job is to transfer his descriptive data onto a keypunch form. Assuming this is a reasonable generalization, an ideal data collection system would then call for the individual to enter his data directly onto a keypunch form (which need not look like the standard 80 column form, but which must be marked with column numbers that are recognizable to the keypunch staff), or, in some cases, for the individual to interact with the computer, via terminal, in a question/answer dialogue. If the questions are easily understood, if they are available in the language spoken by the individual, and if he prints (or types, if a terminal is used) acceptably well, this may indeed be the best way to capture the data.

---

[39] Ibid.

Form 3.2.1-1 Inherent Characteristics of Biographical Data Elements

NON-UNIQUE IDENTIFIER		RECOGNITION FACTOR Is the meaning of the variable easily understood?		RECOLLECTION FACTOR Is the data easy to recall?		CONSISTENCY FACTOR Is the data easy to record accurately?	
		YES	NO	YES	NO	YES	NO
LAST NAME	MALE	✓		✓			✓
	FEMALE	✓		✓			✓
FIRST NAME		✓		✓			✓
MIDDLE NAME		✓		✓			✓
SEX		✓		✓		✓	
BIRTHDATE		✓		✓		✓	
BIRTHPLACE		✓		✓			✓
MARITAL STATUS			✓	✓		✓	
ADDRESS		✓		✓			✓
OCCUPATION			✓	✓			✓
EDUCATION			✓	✓			✓
PHYSICAL CHARACTERISTICS		✓		✓		✓	

Form 3.2.1-2 Data Base Definition Weighting Factors for Biographical Data Elements

NON-UNIQUE IDENTIFIERS		Has this information always been requested in the same way?		Are there instructions for filling out the form with examples showing the intended interpretation?		Are compound names handled? (E.g. are there rules for handling surnames (Von/Van der/ de la, etc.) & place names (Des Moines)?*		Is the system capable of handling very long names?*		Are codes carefully specified? (E.g. under OCCUPATION are categories listed? Under ADDRESS, can suburb names be confused with city names as in Los Angeles 90024 vs. Westwood?)**	
		YES	NO	YES	NO	YES	NO	YES	NO	YES	NO
LAST NAME	MALE										
	FEMALE										
FIRST NAME											
MIDDLE NAME											
SEX											
BIRTHDATE											
BIRTHPLACE											
MARITAL STATUS											
ADDRESS											
OCCUPATION											
EDUCATION											
PHYSICAL CHARACTERISTICS											

*Applicable only to NAME, BIRTHPLACE, and ADDRESS fields.
**Applicable only to BIRTHPLACE, MARITAL STATUS, ADDRESS, OCCUPATION, EDUCATION and PHYSICAL CHARACTERISTICS fields.

The problem is usually not so straightforward, however. If any ambiguity exists in the wording of the question, or if coding of the data is required, it may be far more relevant to have the forms filled out by trained interviewers who understand the intent of the questions and can guide the individual in determining what is for him, the appropriate answer.

These generalities concern data collected specifically for entry into the computer. There is another set of instances which involve collection of data after the fact, as in the case in the Uniform Hospital Discharge Data Demonstration summary. Here existing medical records are being assigned to the abstracting task. The methodology for this abstracting procedure has been carefully laid out. Even so, the accuracy checks carried out by means of a re-abstracting procedure for the data collected in Months 1 and 4 at each hospital in some cases reflect a surprisingly high error rate.

In the following paragraphs, the re-abstracting procedure is quoted as it appears in the project manual, along with the summary rates of errors that occurred.

> "The demonstration design called for two samples of medical records to be re-abstracted in each test site. The first re-abstracting was designed to compare the data produced by applying the UHDDD definitions to the data generated from the same medical records using the definitions in effect prior to the demonstration. The second re-abstracting compared the data produced by the hospital during the fourth month of the demonstration, with the data re-abstracted by the project medical record librarians...Table 2 provides a summary of the discrepancy rates for individual items from the second re-abstracting analyses. In almost all instances, the errors with items having low discrepancy rates were the result of clerical errors; e.g., an error in transcribing data from the medical record to the abstract.

Table 2 Percentage of Abstracts with Discrepancies by Test Sites

ITEM	MAINE	PENN.	WIS.	CALIF.
PERSON I.D.	0.1	0.6	NA*	0.1
DATE OF BIRTH	0.3	1.2	---	0.7
SEX	0.8	0.6	1.2	0.5
MARITAL STATUS	1.0	0.8	1.8	1.1
RACE	0.2	0.4	1.0	0.5
ZIP CODE	---	1.1	2.6	0.5
ADMISSION DATE	0.1	---	0.3	---
ADMISSION HOUR	1.3	24.0	15.5	2.0
DISCHARGE DATE	0.1	0.2	NA	---
ATTENDING PHYSICIAN	0.4	1.4	11.6	4.1
OPERATING PHYSICIAN	0.1	2.4	19.7	NA
PRINCIPAL DIAGNOSES	8.3	11.8	10.7	NA
OTHER DIAGNOSIS	14.6	21.0	30.5	NA
PRINCIPAL PROCEDURE	1.9	6.5	22.4	NA
OTHER PROCEDURES	1.5	9.3	15.8	NA
DATES OF PRINCIPAL AND OTHER PROCEDURES	0.1	7.4	18.2	---
SERVICE TO WHICH ADMITTED	2.6	3.2	3.7	2.5
DISPOSITION OF PATIENT	0.4	1.2	3.9	0.4
PRINCIPAL SOURCE OF PAYMENT	3.7	2.2	1.6	2.7 [40]

*Not Available

Obviously the principal difficulty in this data collection procedure is in the training of clerical staff charged with properly interpreting the available data. C. R. Newing, in discussing data collection problems in the insurance business, points up the necessity of checking data collection procedures and tracing errors back to the point of origin.

> "In our efforts to control quality, training of clerical staff plays a large part particularly as most of the data sheets which we use are quite complicated and contain a large amount of information in coded form. We have found it worth while to return documents which contain errors to the originator and while this is not popular it does seem to be having effect. There was a time when 20 percent or so of documents contained errors but we have reduced this to around 5 percent... Having adopted the principle that the files which we hold are the responsibility of the user departments as far as content and accuracy are concerned, we co-operate in helping them to apply any checks which they may wish to operate."[41]

---

[40] Ibid., p. 38.

[41] Newing, C. R., "Data Arrangement and Checking in the Insurance Business," Applied Statistics, Vol. 2, No. 1, 1972, pp. 45-52.

In another paper on a similar subject, W. F. Kemsley emphasized the detailed analysis required for data that is initially "crude" in form. He reported on a survey directed at collection of "...statistical data on monetary transactions in the personal section." Data collection procedures called for family interviews as well as preparation of a 14-day diary by the family. Many problems were encountered by the editors.

> "Some deficiencies become obvious as soon as the entries are examined; others come to light when data in one part of the budget are compared with another part. Some inconsistencies are found in comparisons between diary entries and interview data; others when interview data are compared with individual or family circumstances."[42]

In general, he felt that the necessity for making editing changes had resulted from:

"(a) duplicated information;
(b) difficulties in applying definitions;
(c) incomplete or inconsistent information;
(d) timing of transactions;
(e) changes in household composition."[43]

Form 3.2.2-1 contains a set of questions relevant to the methodology of data collection for each item. If the information contained in the file was handled differently at different stages of evolution, the analyst may want to fill out two sets of tables, or he may choose to use the latest methodology. If he does this, he must remember that his final results do not fully reflect the status of the current file.

3.2.3 Keying of Data and Input Controls. The entry of data into a computer file is accomplished in one of two ways: interactively, via terminal operating in a real time environment, or batch, with data being keyed ahead of time to cards, tape, disk, etc. Error rates for these methods tend to be in the same general range (from 1-4 percent). Consequently, it is the degree of control placed on the overall handling of the data by the system that makes the difference, and not the choice of keying device. Entry of data via terminal, however, creates an entirely different set of problems. In this case, control is a function of system efficiency and error rates will depend on the degree of sophistication of the programs themselves. James Martin has discussed this difference in approach:

---

[42] Kemsley, W. F., "Pre-Computer Editing of Budgets for the Family Expenditure Survey," Applied Statistics, Vol. 2, No. 1, 1972, pp. 58-64.

[43] Ibid.

Form 3.2.2-1 Data Collection Procedure Weighting Factors for Biographical Data Elements

NON-UNIQUE IDENTIFIER	Was the form filled out by the individual himself or by an interviewer in his presence?		Was data printed or typed? (not handwritten)		Was the original form used as the source for data entry?		If the data was obtained from extant documents, was a re-abstracting check performed?		Is there an on-going training effort for clerical staff covering this item.	
	YES	NO	YES	NO	YES	NO*	YES	NO	YES	NO
LAST NAME										
FIRST NAME										
MIDDLE NAME										
SEX										
BIRTHDATE										
BIRTHPLACE										
MARITAL STATUS										
ADDRESS										
OCCUPATION										
EDUCATION										
PHYSICAL CHARACTERISTICS										

*Enter 1 check mark for each time the data was transferred to a separate form.

> "In some systems all information stored originates from the terminal operators. Sometimes the terminal operators work in a fairly casual manner, compared with the card-punch operators and verifiers of batch data processing. Clearly the viability of the concept of building up a data base in this way depends upon whether we can control the accuracy of the terminal input, catching the errors that occur.
>
> "On certain systems the errors are cumulative. The files contain information about a set of items that is kept for several months and is updated periodically by terminal operator actions. If occasional operator actions cause errors in the files, as the months pass by the files will steadily collect more and more inaccuracies. This situation could clearly bring the system into disrepute, and controls must be devised to prevent it.
>
> "A number of factors make real time systems worse than batch systems for the control of accuracy. First, there are likely to be more terminal operators creating the input and they are scattered over many areas. They tend to be more diverse and less controllable than the operators in a keypunch room. Second, the verification operation found in a keypunch room is usually not employed. Third, batch totals and other batch controls often cannot be used, as the transactions originate singly, not in batches. Fourth, equipment failures will be experienced. It is often when a terminal, line, or computer fails, or during the recover period, that errors originate.
>
> "We do, however, have one factor that is strongly in our favor in real time systems. This is that in an appropriately designed diaglogue most of the errors made can be caught in real time as the operator makes them. The mistake or discrepancy is then rectified on the spot."[44]

Because of the differences in approach between batch and interactive systems, the respective weighting factors are being handled separately. (See Form 3.2.3-1 and 3.2.3-2).

3.2.4 Computer Edit Routines. Checks by computer program of value ranges, formats, data consistency, etc., not only have the advantage of allowing for rapid processing of the data, but further act on data that has already been stored in the computer file and is, therefore, past most of the stages at which new errors are introduced. Susan Wooldridge et al. comments on the importance of this process.

---

[44] Martin, James, op.cit., pp. 70-71

Form 3.2.3-1 Keying and Input Control Weighting Factors for Batch Input

NON-UNIQUE IDENTIFIER		Was data verified as part of the input procedure?		Was input batched and were batch totals checked?		Were check digits used?	
		YES	NO	YES	NO	YES	NO
LAST NAME	MALE						
	FEMALE						
FIRST NAME							
MIDDLE NAME							
SEX							
BIRTHDATE							
BIRTHPLACE							
MARITAL STATUS							
ADDRESS							
OCCUPATION							
EDUCATION							
PHYSICAL CHARACTERISTICS							

Form 3.2.3-2 Keying and Input Control Weighting Factors for Interactive Input

NON-UNIQUE IDENTIFIER		Was data verified as part of the input procedure?		Are adequate backup procedures available?		Do the programs allow the operator to see the data before updating it?		Are there adequate cont over who can change whi data items?	
		YES	NO	YES	NO	YES	NO	YES	NO
LAST NAME	MALE								
	FEMALE								
FIRST NAME									
MIDDLE NAME									
SEX									
BIRTHDATE									
BIRTHPLACE									
MARITAL STATUS									
ADDRESS									
OCCUPATION									
EDUCATION									
PHYSICAL CHARACTERISTICS									

> "The first program in any system should be an edit, or a verification program. All data for a system should go into one simple edit program, if at all possible... The basic principle of the edit program is that it detects all errors that are logically possible to detect, including - or perhaps especially - rare and unlikely ones, for these are the errors which, if allowed to pass into the system cause the most trouble later on."[45]

Programmed checks will, of course, reflect the characteristics of the data in the file but may include format checking, (is there a number in the numeric field?), range (is the age given as greater than 99?), internal consistency (do the birthdate and current age agree?), reasonableness (would someone really have 30 children?), etc. In some cases, an error situation is obvious as in the case of alphabetic data occurring in a numeric field. Where a determination of reasonableness of the data is required, the computer program may not be able to make a final determination. In such case the data must be flagged for further action by the data processing department.

Other important factors in determining the likelihood of errors in the file are the length of time for which the computer editing program has been available as well as consideration of the points in time when improvements have been made in its implementation. If only part of the data has been run against the edit program, a first step in data cleaning may well be to run the entire file through the editing routine to assure that all obvious errors have been caught.

Form 3.2.4-1 lists some of the principal areas of error detection that can be performed by a computer editing program. The analyst is encouraged to check the variables in the current file against this list.

3.2.5 Manual Checks for Accuracy. Valuable as other checks on data validity are, there is no substitute for human judgment in reviewing the records. Final determination of the accuracy of a record demands that the printed version of the data be manually checked against the original document. In describing the creation of new files, Wooldridge, Corder and Johnson state that:

> "...after the data has been transferred to magnetic files the first time, it must be printed out in full. Every item must be checked. This is a time-consuming and tedious job; it is the responsibility of the user department, and plans should be made well in advance for laying on extra staff, authorizing over-time, etc. All discrepancies between the computer files and the manual ones must be checked out. Sometimes the error will be in the old files, in which case it must be corrected from

---

[45] Wooldridge, Susan, et al., op.cit., p. 58.

Form 3.2.4-1 Computer Editing Weighting Factors

NON-UNIQUE IDENTIFIER		Does an adequate edit routine exist for this variable?*		Has all data in the file been run against this edit routine?		Does it check for errors in field formats? (Alpha data in numeric field, etc.)		Is missing data reported?		Are there reasonableness checks? (No married 6-yr. olds etc.)		Are there checks on internal consistency? (Is reported age consistent with birthdate?)	
		YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO
LAST NAME	MALE												
	FEMALE												
FIRST NAME													
MIDDLE NAME													
SEX													
BIRTHDATE													
BIRTHPLACE													
MARITAL STATUS													
ADDRESS													
OCCUPATION													
EDUCATION													
PHYSICAL CHARACTERISTICS													

*If no, place checks in all NO columns on this form for the variables affected.

> scratch... *it is very dangerous to start live processing with unchecked data, and it should never be authorized*. Print-out and checking should continue for the first few cycles of the system; undetected program bugs and input errors are not only likely but *extremely likely*."[46]

Robert Goldstein, in his *Datamation* article entitled "The Cost of Privacy" has suggested that, as standard procedures are adopted for reviewing the contents of files for compliance with the Privacy Act, approximately 10% of the records will either be dropped or rechecked in any given year. Under such a procedure, he estimates that:

> "For a typical system, about 10% of the records would expire each year, and 90% of these would be rechecked rather than dropped."[47]

This emphasizes the importance of such manual checks. Richardson and Cunningham have pointed out that audit trails can be kept with each data item in the file in such a way as to enhance the review of edited data.

> "Reviewing of edited data can benefit considerably from 'data-marking', i.e., storing a bit-string with every item to indicate whether the item:
>
> (i) has been queried without correction;
> (ii) is to be accepted regardless;
> (iii) has been corrected manually;
> (iv) is a candidate for automatic correction later;
> (v) has already been automatically corrected."[48]

An evaluation of the current file must include an analysis of the type of manual checking procedures employed to date. Form 3.2.5-1 gives the reader an opportunity to score the variables in his file with respect to these manual checking procedures.

3.3 SUMMARY OF WEIGHTING FACTORS

Form 3.3-1 can be used to determine final weighting factors for each of the standard biographical data elements in the table. The check marks appearing in the NO columns of each of the preceding charts should be totalled for each variable and the results entered in this form. These results are used to determine candidate fields where error and omissions are likely to occur.

---

[46] *Ibid.*, p. 61.

[47] Goldstein, Robert C., "The Cost of Privacy," *Datamation*, October 1975, p. 65.

[48] Richardson, M. and A. D. Cunningham," Vetting of Industrial Survey Questionnaires,' *Applied Statistics*, Vol. 2, No. 1, 1972, p. 54.

Form 3.2.5-1 Weighting Factors for Manual Accuracy Checks

NON-UNIQUE IDENTIFIER		Is the record printed out after storage in file?		Is the printed output checked against original document?		Are update procedures audited by printing and verifying the updated record?		Did the personnel receive special training for this work?	
		YES	NO	YES	NO	YES	NO	YES	NO
LAST NAME	MALE								
	FEMALE								
FIRST NAME									
MIDDLE NAME									
SEX									
BIRTHDATE									
BIRTHPLACE									
MARITAL STATUS									
ADDRESS									
OCCUPATION									
EDUCATION									
PHYSICAL CHARACTERISTICS									

Form 3.3-1 Cumulative Weighing Factors for Standard Biographical Data Elements

NON-UNIQUE IDENTIFIER		DECAY FACTORS	INHERENT CHARAC-TERISTICS	DATA BASE FACTOR	COLLECTION PROCEDURE FACTOR	INPUT PROCEDURE FACTOR	COMPUTER EDITING FACTOR	MANUAL CHECK FACTOR	CUMULATIVE WEIGHTING FACTOR
LAST NAME	MALE								
	FEMALE								
FIRST NAME									
MIDDLE NAME									
SEX									
BIRTHDATE									
BIRTHPLACE									
MARITAL STATUS									
ADDRESS									
OCCUPATION									
EDUCATION									
PHYSICAL CHARACTERISTICS									

4.0 FILE VALIDATION

4.1 INTRODUCTION

Agencies currently maintaining personal data files will be concerned with the development of procedures for complying with the Privacy Act provisions regarding the relevance, timeliness, and accuracy of data in the individual records. The file validation techniques thus established will also have implications with respect to the evaluation of the level of accuracy maintained in the file over time. Statistics reflecting the number and type of problems detected when selected records are processed should give a clear picture of improvements in data handling. For purposes of the current discussion, performance of file validation is crucial to formulation of a reasonably accurate estimate of the retrieval reliability for each of the variables maintained in the records.

4.2 VALIDATION TECHNIQUES

Section 3 emphasized the importance of the source document with respect to the ultimate accuracy level of the file. The CUMULATIVE WEIGHTING FACTOR column in Table 3.3-1, Cumulative Weight Factors for Standard Biographical Data Elements, gives the total number of negative points accumulated for each variable during that analysis procedure. If all of those variables have approximately the same final scores, then the information gained is not discriminatory, but rather implies what is to be anticipated in the area of accuracy for the entire file. If, however, a few of the variables have much higher scores than the others, these should automatically be eliminated from consideration as retrieval keys. These values are not as directly quantifiable as error and omission rates, but should instead serve as guidelines in the retrieval key selection process and, incidentally, as indicators regarding future use of these variables.

Evaluation of error and omission rates must be accomplished through manual checks of the source documents against listings of the records contained in the computer file. The first decision to be made regarding this procedure involves determination of the number of documents to be checked. A reasonable starting point would be to consider using a small test sample, say 100 records, for this procedure, and then to select a sample size according to the procedure given in Section 5.4. It appears that a determination of error/omission rates would be satisfactory if the standard deviation is less than or equal to .01. The procedure described in Section 5.4 is based on this degree of accuracy.

Once the sample size has been established, the selection of the actual records can be made. The most important factor here is that the records must be selected from different physical locations within the data base. This is essential to avoid skewing of the results due to system problems such as the existence of bad tape, overwritten disk blocks, etc. Further, since both the accuracy and the nature of the data may have varied over time, randomness in selection of records is essential to a fair evaluation of the state of the data base. The number of physical locations used in a given data base will depend upon its size, but that number should not be less than four if a reasonable sampling is to be achieved.

For purposes of recording error and omission counts, a chart similar to that shown in Table 4.2-1 should be created. This chart may also serve as an audit document demonstrating the problems detected, resulting error analysis, problems remaining unresolved, action to be taken, and, finally, indication that the file has been corrected and checked. The analysis results for each record should be kept on a separate sheet for ease of use and also as a precaution against the introduction of confusion between records.

4.3 FILE VALIDATION EXAMPLE

An example of the techniques involved in manual evaluation of the accuracy of the information contained in the files is given in this section. This discussion is based on use of a small dummy data file consisting of one hundred records constructed through random accessing of the telephone book and the random number generator. Separate accesses were made for each component of each name and for each number group. Thus the entries are entirely fictitious .

The ten records (in this case ten percent of the file) selected for validation are shown, in the most accurate source format available, in Table 4.3-1. This data is assumed to exist on the application forms filled out by each of the individuals. It is also assumed for purposes of illustration that the printouts shown in Table 4.3-2 were obtained for each of these records from the computer file.

Table 4.3-3 contains the results of the manual check performed on records number 20 and 30. In this case, data has been stored in the file for a set of twins, HECTOR and HARRY GENERAUX. As is clear from the source documents, the twins lived at the same address at the time the applications were filled out. In the intervening period, HECTOR moved to Laurel Street, and a system update was initiated by the operator when he reported his new address. The update to the record was made on the basis of SURNAME, and BIRTHDATE, and thus it was easy for the mistake to occur.

In record number 70 an obvious case of number transposition has occurred in the address field, in this case affecting both street and house number. Record 80 has a mistake in the spelling of the SURNAME where an 'I' has incorrectly been input as 'E'. And in record 100, a mistake has occurred in entering the birthdate. These errors are summarized in Table 4.3-4.

Use of these rates is discussed in Section 5.

Table 4.2-1 Suggested Format for File Validation Record

RECORD ID ______________ INITIAL CHECK BY ____________ DATE ____________

BIOGRAPHICAL DATA ELEMENT	SOURCE DOCUMENT(S)	ERROR DETECTED		CORRECTED BY (initials) ON (date)	LISTING CHECKED: BY (initials) ON (date)
		MISSING DATA	INCORRECT DATA		

Table 4.3-1 Source Documents for Ten Sample Records from the Dummy Data Base

SOURCE OF DATA	CASE #	SURNAME	FIRST NAME	MIDDLE	STREET ADDRESS	DATE OF BIRTH
Application Form	10	BROADWELL	MEYER	A.	3690 S. Oxford	September 18, 1906
	20	GENERAUX	HECTOR	M.	3872 Tremne	July 7, 1932
	30	GENERAUX	HARRY	M.	3872 Tremne	July 7, 1932
	40	INIQUEZ	VIRGINIA	A.	1409 Avenue Fifty-One	September 18, 1953
	50	LIRA	JAIME	J.	3018 Second Ave.	October 1, 1938
	60	NARITA	MASATO	-	5611 Towne	November 22, 1925
	70	RICHMOND	JOHN	-	7012 W. 41st	December 28, 1909
	80	STRIGLOS	ARNOLD	S.	1513 Beverly	February 4, 1952
	90	WALL	ROY	E.	2800 Geer	May 18, 1936
	100	ZWEIG	EDWARD	-	430 2nd Avenue	October 15, 1955
Update Sheet	20	GENERAUX	HECTOR	M.	4904 N. Laurel	July 7, 1932

Table 4.3-2 Computer Listings for Ten Sample Records from the Dummy Data Base

CASE #	SURNAME	FIRST NAME	MIDDLE INITIAL	STREET ADDRESS	DATE OF BIRTH
10	BROADWELL	MEYER	A.	3690 S. Oxford	9/18/06
20	GENERAUX	HECTOR	M.	3872 Tremne	7/7/32
30	GENERAUX	HARRY	M.	4904 N. Laurel	7/7/32
40	INIQUEZ	VIRGINIA	A.	1409 Avenue Fifty-One	9/18/53
50	LIRA	JAIME	J.	3018 Second Ave.	10/1/38
60	NARITA	MASATO	-	5611 Towne	11/22/25
70	RICHMOND	JOHN	-	7041 W.12th	12/28/09
80	STREGLOS	ARNOLD	S.	1513 Beverly	2/4/52
90	WALL	ROY	E.	2800 Geer	5/18/36
100	ZWEIG	EDWARD	-	430 2nd Avenue	10/10/55

Table 4.3-3 File Validation Record for a Sample Record from the Dummy Data Base

RECORD ID 30

INITIAL CHECK BY GM

DATE 12/10/75

BIOGRAPHICAL DATA ELEMENT		SOURCE DOCUMENT(S) APPLICATION FORM	ERROR DETECTED		CORRECTED BY (initials) ON (date)	LISTING CHECKED: BY (initials) ON (date)
			MISSING DATA	INCORRECT DATA		
SURNAME		X				
FIRST NAME		X				
STREET NUMBER		X		X SHOULD BE 3872		
STREET ADDRESS		X		X SHOULD BE TREMNE		
DATE OF BIRTH		X				

Table 4.3-4 Summary of Errors and Omissions Detected in Analysis of 10% of the File

BIOGRAPHICAL DATA ELEMENT	TOTAL NUMBER OF ERRORS DETECTED	
	MISSING DATA	INCORRECT DATA
SURNAME		1
FIRST NAME		
STREET NUMBER		3*
STREET NAME		3*
DATE OF BIRTH		1

*Since the wrong record was updated for the Generaux twins, and another address error exists, there are a total of 3 incorrect addresses in the file.

5.0 COMPUTER SIMULATION OF DATA BASE RETRIEVAL

5.1 PRECISION AND RECALL

The preceding sections have introduced methodologies for determining the reliability of currently existing data bases. It was seen that a large part of the privacy problem reduces to the technological capabilities of information storage and retrieval systems. This section therefore deals with concepts related to these questions, such as effectiveness of retrieval techniques, precision, and recall. Finally, the quantification of these concepts provides tools for information systems managers to use in evaluation of the effectiveness of their systems. Computer simulations are a powerful aid in performance of such evaluations. Accordingly, a mathematical model and its associated computer simulation are presented in this section.

As noted in Section 2.4, the effectiveness of information storage and retrieval systems has classically been measured by two parameters: recall and precision. Recall is defined as the percentage of relevant records which are retrieved in response to a particular query, while precision is the percentage of the retrieved records which are relevant. An ideal system returns 100% of the relevant records in the data base in response to each query, and, in that subset of the data base retrieved by the system, 100% of the records are relevant. Such high levels of recall and precision can be simultaneously achieved under the following conditions:

1. Queries refer only to key fields where the data is perfectly recorded (no errors or omissions).

2. All of the records in the data base are examined to determine if they meet the query criteria.

Many storage and retrieval systems are operated as if these conditions were true when, in fact, they are known to be false. Condition 1, perfectly valid data, can be approached, but never really reached; condition 2, examination of all records in the data base, can be simulated by the use of indexed file structures and multi-level searches. Indeed, when queries consist of only terms upon which indexes have been built, condition 2 holds for real data bases.

It should be pointed out, however, that the recall and precision figures discussed in the preceding paragraph refer only to the relationship between the query as stated to the system and the retrieved records which resulted from the query. Normally this process is simply a surrogate for the more complicated process of locating the record or records in a data base which match some aspect of the real world. Often the criteria by which the user wishes to select a record are not those criteria which were embodied in the indexing system of the data base, thus causing the user's query to be translated into one or more system-oriented queries before the search is carried out. Under these conditions, the most interesting queries are likely to be those which the system cannot answer, and which are thus not translatable into equivalent system oriented query formulations. Consequently, the system-oriented translations of such queries, even if answered with 100% recall and precision, do not provide high levels of recall and precision with respect to the real query -- that specified by the user.

Errors in the data base become most important when queries are made very specific and when total recall (with respect to the real world, not just to the query) is required. When, for example, there is a statutory requirement to identify the records in a data base which contains personal data in order to provide benefits to those persons who meet the necessary qualifications, there is at least a social cost associated with neglecting to properly identify those individuals who should be provided with the benefits, and there is an economic cost which results from improperly providing benefits to ineligible recipients. There is also a statutory requirement that agencies which maintain personal data bases provide, upon demand, the records which pertain to an individual and allow the subject of such records to challenge errors of fact, etc. The problem here is one of providing the inquirer with his, and only his, record, without requiring him to provide sufficient information to create a record on him if it does not already exist.

Errors in the data base may be of two types: destructive, where the erroneous data in a particular field errs in such a way that it will not satisfy any real query against the data base, and transformational, where the content of a data field is wrong, but satisfiable. An example of the first type of error would be something like AGE=2B (alpha data in a numeric field), which will never match a numeric query statement. An example of a transformational error is NAME=SMITH (when the name is really SMYTHE). Destructive errors cause a record to become inaccessible via the particular key field in which the key value has been destroyed, while transformational errors will cause the record to be returned in response to queries for which it is irrelevant, and to be missed when it should be found. The effect of error rates in the various data fields employed in a query can be quantified as follows:

1. Given a query Q consisting of the conjunction of n keys K_1, K_2, K_3, ... K_n, where the K_i each represent a particular value of a particular key field in the data base records, and

2. If each of the fields in the query Q is associated with confidence factors (equal to 1 minus the sum of error and omission rates) c_1, c_2, ... c_n.

then the recall ratio associated with the response to query Q is the product of the confidence factors associated with each of the key fields:*

3. $G = \prod_{i=1}^{n} c_i$

It follows that queries which reference more key fields in an attempt to achieve more precision tend to be subject to smaller recall ratios since the c_i are always less than or equal to 1.

5.2 QUERY FORMULATION FOR BEST RESULTS

Queries should be formulated using the most restrictive key fields in order to provide the most specific answers. Additional precision can be achieved by utilizing the conjunction of more fields in each query, although this may result in a decline in the recall ratio, thus causing records to be missed which are actually contained in the data base. Equally, those fields with the lowest error rates (i.e., with the highest confidence factors) should be preferred over those known to be badly in error. Qualitatively, the following order provides the greatest discrimination among records:

1. Name (Surname and Given Name), although the discriminatory power of this field varies depending on the nature of the surname. Matching a name like TREFFTZS provides much more information than matching a name like SMITH, for example.

2. Surname.

3. Residence address (Street Number, Street Name, City, and State), although this information tends to decay rapidly with time.

---

*Recall ratio is defined as the proportion of relevant records retrieved in response to a query, or equivalently, as the probability of a relevant record being retrieved. For the purpose of this discussion (as pointed out above), a relevant record is one corresponding to an individual who has exactly those attributes defined by the query. However, the actual recorded key values may either be in error or omitted. Thus, the probability of the key value in the ith key field being correct, and hence retrievable by the query, is c_i. If the retrieval mechanism is such that a record is retrieved only if it matches on each of n key values in a query, then it follows that the probability of a relevant record being retrieved is the product of all the n probabilities; i.e.,

$$G = c_1 \times c_2 \times \cdots \times c_n = \prod_{i=1}^{n} c_i$$

and so this is the recall ratio.

81

4. Date of Birth (Year, Month, and Day).

5. Place of Birth (City, State), although this also varies: Sparta, Illinois as a birthplace provides more identifying information than does New York, New York.

6. Sex. This is an example of a key field which provides very little information -- it may partition the file into two equal halves at best. In more extreme cases where the contents of the file are heavily skewed (all males or all females) this field provides, in general, even less information, only being useful in the rarer cases.

Subjective estimates on the discrimination of various keys are provided in Table 5.3-1 while statistical data on surname records are given in Table 5.3-2.

5.3 SOFT MATCH TECHNIQUES

"Soft-Match" techniques may be utilized to overcome the decline in recall which is normally associated with improvements in precision. The increased error rate which is associated with increases in the number of variables entering into the query can be reduced by examining the records retrieved by the sub-queries. For example, a query containing three key fields would ideally return a record (or set of records) in which the three required values are found. Errors in the recording of either the original data or the query may, however, have produced only a partial match, i.e., for either one or two keys. The problem is to devise a method for accounting for these partial matches. Such methods are called "Soft-Match" techniques.

The result of a Soft-Match Technique is not only a set of retrieval records, but also a numerical weight assigned to each record to indicate its degree of responsiveness to the query. In Appendix III a specific weighting scheme for a Soft-Match Technique is presented.

5.4 EXAMPLE OF USE OF FILE VALIDATION SUMMARY DATA

In Section 4.3, an example of the use of file validation techniques is given. Tables 4.3-4 detailed the summary of error and omissions within the sample data file. The final tally of errors shown in the table for each of the variables is:

Variable	Error Count Per 10
SURNAME	1
FIRST NAME	0
STREET NUMBER	3
STREET NAME	3
DATE OF BIRTH	1

Table 5.3-1 Retrieval Ratios (Estimates of Proportion of File Retrieved)*

Field Number	Contents	Retrieval Ratio
1	Day and Month of Birth	1/365
2	Month of Birth	1/12
3	Year of Birth	1/50
4	State	1/50
5	Sex	1/2
6	Street Number	1/10,000
7	Street Name	1/5,000
8	Education	1/15
9	County	1/3,000
10	Date of Birth (YMD)	1/18,000

*The values used are based on subjective estimates taking into consideration the meaning of the keys. These values are used in the model described in Section 5.5.

Table 5.3-2 Statistical Classification of Surname Records as compiled from Social Security Administration Accounts.[49]

DECILE*	Number of Records Retrieved**	% of File Retrieved**
1	16.35	.00000987%
2	408.17	.0002459%
3	1,599.28	.0009635%
4	4,515.456	.00272037%
5	8,000.00	.00481966%
6	10,999.94	.006627%
7	20,997.32	.01265%
8	80,000.62	.048197%
9	130,000.80	.07832%
10	561,773.3	.3384%

*Decile 1 corresponds to the least common names, Decile 10, the most common (e.g., Smith).

**Based on 1 value per surname key. Each decile contains 10% of the population.

---

[49] "Report of Distributions of Surnames in the Social Account Number File," Department of Health, Education and Welfare, June 1, 1964.

These error rates imply that the file as a whole is heavily riddled with serious errors: the single error in the surname sample implies that 10 of the records in the file are coded with an erroneous SURNAME -- this is 10% of the file. Three detected errors in the STREET NUMBER again imply that 30% of the street numbers in the file as a whole are in error, etc. Based on this information, a typical query against this file, using the keys of SURNAME, GIVEN NAME, STREET NUMBER, STREET NAME, and DATE OF BIRTH, would have a combined confidence/recall factors of only 39.69% -- less than half of the truly relevant records would ever be located in response to a query.

Normally, of course, statistical methods would not be used on a file as small as the dummy data base discussed here, but would be used for a file containing thousands of records. The procedure is as follows:

Suppose we want to determine the error rate (or the omission rate -- the theory is the same) so that the true value differs from the measured value by less than 0.01 with a high probability (say 66%). This means we want to choose a sample size so that the standard deviation in the determination of the error rate is less than 0.01. The standard deviation is given by the formula:[50]

$$\sigma_{\text{error rate}} = \sqrt{\frac{\text{error rate} \times (1 - \text{error rate})}{\text{Sample Size}}}\ .$$

Therefore, if we wish the standard deviation to be less than 0.01, we must have

$$\text{Sample size} > 10{,}000 \times \text{error rate} \times (1 - \text{error rate}).$$

Since the product of the error rate and its complement is bounded by 0.25, a sample size of 2,500 will assure a standard deviation of less than 0.01. But a smaller sample size may be used, depending on the magnitude of the error rate. To determine this, select a small test sample, say of 100 records. Estimate the error rate for this test sample. Then a safe determination of an adequate sample size is given by

$$\text{Sample size} = 10{,}000 \times \text{estimate of error rate}.$$

For example, if the estimate is 0.1, then the sample size should be 1000.

Having determined the error and omission rates, the confidence factors (see Section 5.1) can be calculated. The recall ratio for any query involving the key fields for which these confidence factors have been developed is simply the product of the confidence factors.

---

[50] Hoel, P. G., Introduction to Mathematical Statistics, John Wiley and Sons, New York, 1954, pp. 87-88, also pp. 107-108. This is the usual formula for the standard deviation of a mean value.

Assuming that the error and omission rates for the variables SURNAME, STREET NUMBER, STREET NAME, and DATE OF BIRTH were 0.01, 0.03, 0.03, 0.01, respectively, then the resulting confidence factors would be:

SURNAME	0.99
STREET NUMBER	0.97
STREET NAME	0.97
DATE OF BIRTH	0.99

and a query involving all four of these fields would tend to retrieve $0.99 \times 0.97 \times 0.97 \times 0.99 = 0.922$ or 92.2% of the relevant records.

Use of these data will be shown below.

5.5 PRECISION AND PROBABLE MATCH

Given a certain retrieved record, the question arises: Is it in fact the correct record for the requester? The answer to this must be expressed as a probability. The theoretical formulation of this probability problem is presented in Appendix II.

A computer program was written to calculate probability values resulting from the use of various combinations of keys. Recall that the basic problem is to evaluate the probability that a record will match a requester given the presence in the record of a number of key values $K_1, \ldots, K_n$ which match the request. The probability values are developed in terms of the following auxiliary quantities.

(1) p = the a priori probability that a randomly selected record corresponds to that record desired by a requester.

(Notation in Appendix II, Equation (16), $p = P(M_i)$)

The range of p should be from $1/T$ (where T is the largest population from which the data base is selected) to $1/N$ (where N is the number of records in the data base).

For each key j on which the search is conducted, let

(2) a_j = error rate in key j (a proportion).

(3) b_j = relative frequency of omission of data in key j.

From (2) and (3), we calculate

(4) $c_j = 1 - a_j - b_j$.

Thus c_j is the probability

$$P(K_j|M_i)$$

given in Appendix II, Equation (8b).

(5) Let n_j be the number of records with specific matching data in key j, i.e., the number of records retrieved matching the requested value of key j. Let N be the number of records in the data base, then define

$$r_j = n_j/N$$

Thus r_j is the probability $P(K_j)$ given in Appendix II, p. 176.

(6) For each key value j we define

$$s_j = c_j/r_j.$$

(7) We then have from Appendix II, Equation (18), the probability $P(M_i|K_j)$ given by

$$q_j = ps_j$$

(8) An auxiliary quantity F (Appendix II, Equation (23)) is then calculated:

$$F = \left(\frac{p}{1 - p}\right)^{n - 1} \prod_{j = 1}^{n} \left(\frac{1 - q_j}{q_j}\right)$$

where n is the number of key values which match the request. The quantity F is used as shown below.

(9) Finally, the required probability (i.e., the probability that a record is that desired by a requester conditional on a match on n key values in the query) is given by

$$V = \frac{1}{1 + EF}$$

where E is a quantization of a certain independence relation between the keys. (see Appendix II, pp. 179-180). The smaller the value of E, the greater the ability of the keys to discriminate. Independent keys mean that E = 1. If $E < 1$, then $V > 1/(1 + F)$.

The probability value V corresponds to the precision of a retrieval of one record.

Results of the computer simulation are shown in Appendix I. The tables should be read as follows:

(1) Each table is identified by the data fields used in the simulated query. All simulated queries utilized the surname field.

(2) The surname decile identifies the discriminatory power of the particular surname used in the query: Surnames in the first decile are the most specific (TREFFTZS, for example) while those in the tenth decile are the more common surnames, such as JONES and SMITH.

(3) The column headings reflect the three sample file sizes which were simulated: 1 million, 4 million, and 220 million records.

(4) The left column of the table reflects the recall factor* obtained by multiplying together the simulated error rate for each of the fields utilized in the query. These rates were varied in parallel from 0 (no errors) to 6% errors in increments of 0.5%. All fields were assumed to be subject to the same error.

(5) The body of each table tabulates the precision of each retrieved result with respect to the query. Values of 1.0 imply that unique records were properly identified. Values less than 1.0 imply that the probability of unique identification is less than 100%.

The model reflects the precision and recall resulting from the attempt to locate the record pertaining to a specific individual in a data base of the specified size, subject to given error rates, for several types of hypothetical queries. Three file sizes are postulated: a file of 220 million records, one of 4 million records, and a smaller file containing only 1 million records. All the queries are assumed to consist of a surname and one or more other key values. The discriminatory power of a query varies with the discriminatory power of the key and key values used and the confidence factor in the data base. The precision of a retrieval made against the three file sizes given the retrieval confidence factor of 92.2% in the example of Section 5.4, and utilizing the four key fields of surname, street number, street name, and date of birth can be read from the fifth line (recall 0.922) of Tables 5a - 5j in Appendix I. The tables present 10 different values for the retrieval precision of such a query, depending on the value of the surname: names such as Jones have far less discriminatory power than names such as Trefftzs or Krcmar. It can be seen that the precision of a query improves as the file size decreases. Files smaller than 1 million records will provide proportionately better results.

---

*The definition of recall is given in Section 5.2.

6.0 CONCLUSIONS AND RECOMMENDATIONS

6.1 CONCLUSIONS

The preceding sections have addressed the problem of accessing individual records from personal data files using non-unique identifiers where accessing includes updating, purging, and other bookkeeping tasks. The value of the legislation requiring Federal agencies to make such records available on request to individuals whose dossiers exist in their files is obvious, but compliance with this legislation is a complex matter for which there can be no single generalized approach. Each agency must determine independently the most appropriate solution given its own operational environment.

Exphasis in this document has been on one aspect of the required implementation: the retrieval of the individual's record through use of identifiers that he can readily recall and that do not include his Social Security account number. The analysis has included examination of the characteristics of standard identifiers with emphasis on the varying selectivity of these variables. The high discriminatory value of the NAME variable resulted in inclusion of descriptions and evaluations of several of the currently available name lookup algorithms. Techniques applicable to this type of retrieval problem were discussed along with indexing schemes ranging from the most straightforward to complex systems of encoded pointers.

The selected examples for generating the precision tables from the mathematical model (see App. I) clearly show that retrieval through use of non-unique identifiers is, indeed, possible. It must be understood, however, that the search keys used in these examples are illustrative only. In practical cases it may be necessary to employ search keys with more discriminating power. Furthermore, such retrieval has been carried out in a variety of ways and has proved to be practical. For example, the 3-level index scheme developed by the Immigration and Naturalization Service presently retrieves 5,000,000 records (anticipated growth up to 10,000,000 records) with a response time per record measured in seconds.

The discussion of available techniques also made it clear that the degree of success resulting from use of such techniques depends directly on the quality of the data items being used as search keys. (The assumption is made that the query data itself if correct and that it has been entered properly.) The reliability of retrieval through use of non-unique identifiers was further demonstrated through use of a probability model programmed to generate the precision (i.e., the probability that a record if retrieved as matching the query, as formulated, on a variety of keys, will be in fact the correct record), given a wide range of recall factors (i.e., the probability that the record, if in the file, will be retrieved based on a range of confidence factors simulating a reasonable level of error and omission rates). Results from this modeling procedure again indicate that retrieval mechanisms based on high discrimination keys are practical.

It is also clear, however, that these techniques must be implemented with great care. The use of non-unique identifiers results in very high precision but rather low recall. This effect is obvious from consideration of the fact that the more identifiers used, the more complete the identification of the individual becomes. At the same time, the introduction of more keys increases the potential for problems in matching exactly the information that is contained in the file. For this reason, a Soft-Match Technique, such as that described in Section 5, may be necessary as back up to the retrieval procedure.

6.2 RECOMMENDATIONS

The Office of Management and Budget has interpreted the language of the Privacy Act to mean that the mere existence of personal data items in a file is not sufficient to bring it under these legislative constraints. In accordance with the definition of "system of records" (Privacy Act, Sec. 552a (a) (5)) these items become relevant only when they are used as retrieval keys:

> "This language further suggests that the Congress did not intend to require that an individual be given access to information which the agency does not retrieve by reference to his or her name or some other identifying particular."[51]

A first step in compliance will then be to determine the exact keys currently in use and from that list to determine whether or not the legislation applies.

The current considerations apply only to the specific problem of the retrieval of an individual's record. But even this single aspect of the legislation will require a review of existing systems to make the following determinations:

1. Are the personal identifiers contained in the file necessary?
2. What non-unique identifiers are currently available as retrieval keys?
3. Are these currently available retrieval keys adequate for accessing the required records in a cost effective way?
4. If not, how can the existing capabilities be expanded or modified to make the required records accessible?

Each of these considerations is discussed below.

---

[51] Office of Management and Budget Circular A-108 and accompanying "Guidelines for Implementing Section 3 of the Privacy Act of 1974," Federal Register, 9 July 1975, Vol. 40, No. 32, pp. 28948-28978.

6.2.1 Evaluation of Current File Access Capability. If it does apply, the agency should make a determination of acceptability of the current retrieval capability in the context of the legislation and the OMB Guidelines. Using the search keys that are both available and relevant, weighting factors should be formed as outlined in Section 3. This procedure should be followed by an analysis of the error and omission rates for these variables as suggested in Section 4. From these rates, confidence factors can be formed, using the methodology described in Section 5, and a determination of precision and recall made for some combinations of these variables using the computer-generated tables shown in Appendix I. If the variables that are already available as retrieval keys are not adequate to the task, then current system capabilities will require expansion. Some discussion of available techniques is given in Section 2 for consideration under these circumstances. It should be noted, however, that the Office of Management and Budget has stressed use of existing capabilities whenever possible.

> "...the development of new retrieval and indexing capabilities is not encouraged, rather agencies should exploit existing capabilities to serve individual needs."[52]

6.2.2 Evaluation of the Data Base. The most difficult aspect of the analysis procedure is that involving evaluation of the data base itself.

An orderly sequence of steps should be taken to evaluate any data base in order to bring it into compliance with the requirements of the Privacy Act. A similar sequence should be followed when establishing any new data bases containing personal data. For existing data bases, the required actions fall into the categories of:

1. Analysis of present data validity,
2. Correction of errors,
3. Retirement of obsolete data, and
4. Establishment of a review cycle to keep data current.

For new data bases, the first of these steps is replaced by an analysis of data requirements and initial data collection steps.

While it is unrealistic to assume that any data base is error-free in any data field, all fields which comprise the data base should contain data which are as accurate as possible. Error rates which exceed four or five percent tend to throw the utility of the data base into question, and fields with such error rates should be either corrected (at some cost) or deleted from the file. The basic test of any personal data base is whether the people represented by the records in the data base can, in fact, be found and shown to be the individuals to whom the records are assumed to refer. As time passes, addresses change, names change, ages change, occupations change, and people move.

---

[52] Loc.Cit.

Files which have not been subjected to ongoing verification and update activities, particularly those in which the subjects are not actively involved in the maintenance process, may be found to contain a large cast of fictional characters.

The requirement for the maintenance of such files should be reviewed at the highest levels--certainly such information should not be made the basis of substantive policy decisions, and the very need for the file in question should be carefully considered. Files which are really necessary, however, should be evaluated periodically to determine their accuracy and to correct erroneous data.

The utility of such files--and compliance with privacy legislation--requires that a high level of accuracy be maintained, and that suitable resources be expended to insure such a level of accuracy.

APPENDIX I

PRECISION TABLES

Table IDs	Variables Selected	Pages
1a - 1j	Surname and Sex	74 - 83
2a - 2j	Surname and Street Number	84 - 93
3a - 3j	Surname and Birthdate (month/day)	94 - 103
4a - 4j	Surname and Birthdate (year/month/day)	104 - 113
5a - 5j	Surname, Birthdate (year/month/day), Street Name and Street Number	114 - 123
6a - 6j	Surname and State	124 - 133
7a - 7j	Surname, Sex, and Birthdate (month/day)	134 - 143
8a - 8j	Surname, Sex, and Birthdate (year/month/day)	144 - 153
9a - 9j	Surname, Street Name, Street Number, and State	154 - 163
10a - 10j	Surname, Sex, Street Name, Street Number, and Birthdate (year/month/day)	164 - 173

Table 1a. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
AND SURNAME

SURNAME DECILE 1

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MILLION
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.88051399E-01
0.50%	0.9900	0.10000000E 01	0.10000000E 01	0.87230503E-01
1.00%	0.9801	0.10000000E 01	0.10000000E 01	0.86412558E-01
1.50%	0.9702	0.10000000E 01	0.10000000E 01	0.85597557E-01
2.00%	0.9604	0.10000000E 01	0.10000000E 01	0.84785554E-01
2.50%	0.9506	0.10000000E 01	0.10000000E 01	0.83976528E-01
3.00%	0.9409	0.10000000E 01	0.10000000E 01	0.83170526E-01
3.50%	0.9312	0.10000000E 01	0.10000000E 01	0.82367547E-01
4.00%	0.9216	0.10000000E 01	0.10000000E 01	0.81567616E-01
4.50%	0.9120	0.10000000E 01	0.10000000E 01	0.80770735E-01
5.00%	0.9025	0.10000000E 01	0.10000000E 01	0.79976950E-01
5.50%	0.8930	0.10000000E 01	0.10000000E 01	0.79186244E-01
6.00%	0.8836	0.10000000E 01	0.10000000E 01	0.78399552E-01

Table 1b. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
AND SURNAME

SURNAME DECILE 2				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220
0.00%	1.0000	0.57820207E 00	0.18457000E 00	0.369
0.50%	0.9900	0.57491933E 00	0.18298137E 00	0.365
1.00%	0.9801	0.57162050E 00	0.18139592E 00	0.361
1.50%	0.9702	0.56830560E 00	0.17981370E 00	0.358
2.00%	0.9604	0.56497469E 00	0.17823478E 00	0.354
2.50%	0.9506	0.56162793E 00	0.17665921E 00	0.350
3.00%	0.9409	0.55826540E 00	0.17508704E 00	0.347
3.50%	0.9312	0.55488715E 00	0.17351835E 00	0.343
4.00%	0.9216	0.55149332E 00	0.17195320E 00	0.340
4.50%	0.9120	0.54808400E 00	0.17039160E 00	0.336
5.00%	0.9025	0.54465935E 00	0.16883367E 00	0.333
5.50%	0.8930	0.54121939E 00	0.16727941E 00	0.329
6.00%	0.8836	0.53776428E 00	0.16572891E 00	0.326

```
ON OF RETRIEVED RESULTS FOR DATA FIELDS
X
D SURNAME
------------------------------------------------------------------------------------
        SURNAME DECILE  2                                                          !
------------------------------------------------------------------------------------
                             FILE SIZE                                             !
------------------------------------------------------------------------------------
    1 MILLION          !            4 MILLION           !         220 MILLION      !
------------------------------------------------------------------------------------
                       !                                !                          !
  0.57820207E 00       !          0.18457000E 00        !       0.36901724E-02     !
  0.57491933E 00       !          0.18298137E 00        !       0.36534642E-02     !
  0.57162050E 00       !          0.18139592E 00        !       0.36169373E-02     !
  0.56830560E 00       !          0.17981370E 00        !       0.35805922E-02     !
  0.56497469E 00       !          0.17823478E 00        !       0.35444293E-02     !
  0.56162793E 00       !          0.17665921E 00        !       0.35084401E-02     !
  0.55826540E 00       !          0.17508704E 00        !       0.34726487E-02     !
  0.55488715E 00       !          0.17351835E 00        !       0.34370318E-02     !
  0.55149332E 00       !          0.17195320E 00        !       0.34015963E-02     !
  0.54808400E 00       !          0.17039160E 00        !       0.33663432E-02     !
  0.54465935E 00       !          0.16883367E 00        !       0.33312722E-02     !
  0.54121939E 00       !          0.16727941E 00        !       0.32963833E-02     !
  0.53776428E 00       !          0.16572891E 00        !       0.32616766E-02     !
------------------------------------------------------------------------------------
```

Table 1c. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
AND SURNAME

SURNAME DECILE 3				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.18805843E 00	0.50581697E-01	0.9430
0.50%	0.9900	0.18644465E 00	0.50096085E-01	0.9336
1.00%	0.9801	0.18483400E 00	0.49612513E-01	0.9243
1.50%	0.9702	0.18322655E 00	0.49130986E-01	0.9150
2.00%	0.9604	0.18162232E 00	0.48651503E-01	0.9057
2.50%	0.9506	0.18002145E 00	0.48174085E-01	0.8965
3.00%	0.9409	0.17842393E 00	0.47698719E-01	0.8873
3.50%	0.9312	0.17692995E 00	0.47225427E-01	0.8782
4.00%	0.9216	0.17523927E 00	0.46754200E-01	0.869
4.50%	0.9120	0.17365220E 00	0.46285061E-01	0.860
5.00%	0.9025	0.17206878E 00	0.45818008E-01	0.851
5.50%	0.8930	0.17048901E 00	0.45353040E-01	0.842
6.00%	0.8836	0.16891298E 00	0.44890170E-01	0.833

```
RETRIEVED RESULTS FOR DATA FIELDS
NAME
------------------------------------------------------------------------------------------
                                                                                         !
 SURNAME DECILE  3                                                                       !
-----------------------------------------------------------------------------------------
                        FILE SIZE                                                        !
-----------------------------------------------------------------------------------------
1 MILLION            !           4 MILLION            !          220 MILLION             !
-----------------------------------------------------------------------------------------
                     !                                !                                  !
18805843E 00         !        0.50581697E-01          !        0.94308455E-03            !
.18644465E 00        !        0.50096085E-01          !        0.93368396E-03            !
.18483400E 00        !        0.49612513E-01          !        0.92433031E-03            !
.18322655E 00        !        0.49130986E-01          !        0.91502349E-03            !
.18162232E 00        !        0.48651503E-01          !        0.90576377E-03            !
.18002145E 00        !        0.48174085E-01          !        0.89655103E-03            !
.17842393E 00        !        0.47698719E-01          !        0.88738526E-03            !
.17682985E 00        !        0.47225427E-01          !        0.87826650E-03            !
.17523927E 00        !        0.46754208E-01          !        0.86919469E-03            !
.17365220E 00        !        0.46285061E-01          !        0.86016991E-03            !
.17206878E 00        !        0.45818008E-01          !        0.85119210E-03            !
.17048901E 00        !        0.45353040E-01          !        0.84226140E-03            !
.16891298E 00        !        0.44890170E-01          !        0.83337749E-03            !
-----------------------------------------------------------------------------------------
```

Table 1d. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
AND SURNAME

SURNAME DECILE 4

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.70912752E-01	0.18212487E-01	0.334
0.50%	0.9900	0.70242635E-01	0.18033273E-01	0.3307
1.00%	0.9801	0.69575102E-01	0.17854903E-01	0.3274
1.50%	0.9702	0.68910182E-01	0.17677383E-01	0.324
2.00%	0.9604	0.68247861E-01	0.17500706E-01	0.3208
2.50%	0.9506	0.67588175E-01	0.17324881E-01	0.3176
3.00%	0.9409	0.66931119E-01	0.17149903E-01	0.3143
3.50%	0.9312	0.66276714E-01	0.16975777E-01	0.311
4.00%	0.9216	0.65624970E-01	0.16802502E-01	0.307
4.50%	0.9120	0.64975889E-01	0.16630076E-01	0.304
5.00%	0.9025	0.64329493E-01	0.16458503E-01	0.301
5.50%	0.8930	0.63685784E-01	0.16287783E-01	0.298
6.00%	0.8836	0.63044777E-01	0.16117917E-01	0.295

N OF RETRIEVED RESULTS FOR DATA FIELDS

SURNAME

SURNAME DECILE 4

FILE SIZE		
1 MILLION	4 MILLION	220 MILLION
0.70912752E-01	0.18212487E-01	0.33412327E-03
0.70242635E-01	0.18033273E-01	0.33079124E-03
0.69575102E-01	0.17854903E-01	0.32747587E-03
0.68910182E-01	0.17677383E-01	0.32417716E-03
0.68247861E-01	0.17500706E-01	0.32089516E-03
0.67588175E-01	0.17324881E-01	0.31762986E-03
0.66931119E-01	0.17149903E-01	0.31438124E-03
0.66276714E-01	0.16975777E-01	0.31114931E-03
0.65624970E-01	0.16802502E-01	0.30793401E-03
0.64975889E-01	0.16630076E-01	0.30473544E-03
0.64329493E-01	0.16458503E-01	0.30155356E-03
0.63685784E-01	0.16287783E-01	0.29838939E-03
0.63044777E-01	0.16117917E-01	0.29523984E-03

Table 1e. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS SEX AND SURNAME

SURNAME DECILE 5

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.40653255E-01	0.10320643E-01	0.1886
0.50%	0.9900	0.40259974E-01	0.10218484E-01	0.1867
1.00%	0.9801	0.39868403E-01	0.10116818E-01	0.1848
1.50%	0.9702	0.39478556E-01	0.10015650E-01	0.1829
2.00%	0.9604	0.39090419E-01	0.99149744E-02	0.1811
2.50%	0.9506	0.38704021E-01	0.98147979E-02	0.1792
3.00%	0.9409	0.38319346E-01	0.97151154E-02	0.1774
3.50%	0.9312	0.37936406E-01	0.96159305E-02	0.175
4.00%	0.9216	0.37555203E-01	0.95172422E-02	0.173
4.50%	0.9120	0.37175740E-01	0.94190502E-02	0.172
5.00%	0.9025	0.36798021E-01	0.93213565E-02	0.170
5.50%	0.8930	0.36422051E-01	0.92241597E-02	0.169
6.00%	0.8836	0.36047829E-01	0.91274603E-02	0.166

RECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
AND SURNAME

SURNAME DECILE 5

DR	FILE SIZE		
	1 MILLION	4 MILLION	220 MILLION
	0.40653255E-01	0.10320643E-01	0.18860356E-03
	0.40259974E-01	0.10218484E-01	0.18672250E-03
	0.39868403E-01	0.10116818E-01	0.18485088E-03
	0.39478556E-01	0.10015650E-01	0.18298866E-03
	0.39090419E-01	0.99149744E-02	0.18113586E-03
	0.38704021E-01	0.98147979E-02	0.17929250E-03
	0.38319346E-01	0.97151154E-02	0.17745858E-03
	0.37936406E-01	0.96159305E-02	0.17563405E-03
	0.37555203E-01	0.95172422E-02	0.17381894E-03
	0.37175740E-01	0.94190502E-02	0.17201328E-03
	0.36798021E-01	0.93213565E-02	0.17021705E-03
	0.36422051E-01	0.92241597E-02	0.16843023E-03
	0.36047829E-01	0.91274603E-02	0.16665283E-03

Table 1f. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
AND SURNAME

SURNAME DECILE 6

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.29730966E-01	0.75165338E-02	0.1371
0.50%	0.9900	0.29440941E-01	0.74419786E-02	0.1358
1.00%	0.9801	0.29152233E-01	0.73677843E-02	0.1344
1.50%	0.9702	0.28864346E-01	0.72939548E-02	0.1330
2.00%	0.9604	0.28578769E-01	0.72204903E-02	0.1317
2.50%	0.9506	0.28294017E-01	0.71473905E-02	0.1303
3.00%	0.9409	0.28010589E-01	0.70746557E-02	0.1290
3.50%	0.9312	0.27728408E-01	0.70022870E-02	0.1277
4.00%	0.9216	0.27447715E-01	0.69302845E-02	0.1264
4.50%	0.9120	0.27168273E-01	0.68586473E-02	0.1251
5.00%	0.9025	0.26890161E-01	0.67873756E-02	0.1237
5.50%	0.8930	0.26613382E-01	0.67164700E-02	0.1224
6.00%	0.8836	0.26337936E-01	0.66459302E-02	0.121

ION OF RETRIEVED RESULTS FOR DATA FIELDS
EX
ND SURNAME

SURNAME DECILE 6

FILE SIZE		
1 MILLION	4 MILLION	220 MILLION
0.29730966E-01	0.75165388E-02	0.13717643E-03
0.29440941E-01	0.74419786E-02	0.13580229E-03
0.29152233E-01	0.73677843E-02	0.13444103E-03
0.28864346E-01	0.72939548E-02	0.13308650E-03
0.28578769E-01	0.72204903E-02	0.13173902E-03
0.28294017E-01	0.71473905E-02	0.13039831E-03
0.28010589E-01	0.70746557E-02	0.12906445E-03
0.27728408E-01	0.70022870E-02	0.12773745E-03
0.27447715E-01	0.69302845E-02	0.12641728E-03
0.27168273E-01	0.68586473E-02	0.12510393E-03
0.26890161E-01	0.67873756E-02	0.12379755E-03
0.26613382E-01	0.67164700E-02	0.12249798E-03
0.26337936E-01	0.66459302E-02	0.12120524E-03

Table 1g. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
AND SURNAME

SURNAME DECILE 7

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.15686289E-01	0.39447738E-02	0.71862
0.50%	0.9900	0.15531641E-01	0.39055400E-02	0.71145
1.00%	0.9801	0.15377724E-01	0.38664996E-02	0.70432
1.50%	0.9702	0.15224551E-01	0.38276540E-02	0.69722
2.00%	0.9604	0.15072109E-01	0.37890030E-02	0.69016
2.50%	0.9506	0.14920410E-01	0.37505466E-02	0.68314
3.00%	0.9409	0.14769446E-01	0.37122833E-02	0.67615
3.50%	0.9312	0.14619223E-01	0.36742153E-02	0.6692
4.00%	0.9216	0.14469740E-01	0.36363417E-02	0.6622
4.50%	0.9120	0.14320997E-01	0.35986621E-02	0.6554
5.00%	0.9025	0.14172994E-01	0.35611773E-02	0.6485
5.50%	0.8930	0.14025734E-01	0.35238867E-02	0.6417
6.00%	0.8836	0.13879215E-0[illegible]	0.34867911E-02	0.6349

```
SION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
AND SURNAME
------------------------------------------------------------------------------
                                                                             !
        SURNAME DECILE  7                                                    !
------------------------------------------------------------------------------
!                                FILE SIZE                                   !
------------------------------------------------------------------------------
!        1 MILLION          !        4 MILLION          !    220 MILLION     !
------------------------------------------------------------------------------
!                           !                           !                    !
!     0.15686289E-01        !     0.39447739E-02        !   0.71862298E-04   !
!                           !                           !                    !
!     0.15531641E-01        !     0.39055400E-02        !   0.71145512E-04   !
!                           !                           !                    !
!     0.15377724E-01        !     0.38664996E-02        !   0.70432318E-04   !
!                           !                           !                    !
!     0.15224551E-01        !     0.38276540E-02        !   0.69722711E-04   !
!                           !                           !                    !
!     0.15072109E-01        !     0.37890030E-02        !   0.69016705E-04   !
!                           !                           !                    !
!     0.14920410E-01        !     0.37505466E-02        !   0.68314274E-04   !
!                           !                           !                    !
!     0.14769446E-01        !     0.37122833E-02        !   0.67615452E-04   !
!                           !                           !                    !
!     0.14619223E-01        !     0.36742153E-02        !   0.66920218E-04   !
!                           !                           !                    !
!     0.14469740E-01        !     0.36363417E-02        !   0.66228574E-04   !
!                           !                           !                    !
!     0.14320997E-01        !     0.35986621E-02        !   0.65540521E-04   !
!                           !                           !                    !
!     0.14172994E-01        !     0.35611773E-02        !   0.64856065E-04   !
!                           !                           !                    !
!     0.14025734E-01        !     0.35238867E-02        !   0.64175201E-04   !
!                           !                           !                    !
!     0.13879215E-01        !     0.34867911E-02        !   0.63497925E-04   !
------------------------------------------------------------------------------
```

Table 1h. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
AND SURNAME

		SURNAME DECILE 8		
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.41410479E-02	0.10368714E-02	0.1886
0.50%	0.9900	0.40998683E-02	0.10265366E-02	0.1867
1.00%	0.9801	0.40588915E-02	0.10162533E-02	0.1848
1.50%	0.9702	0.40181193E-02	0.10060218E-02	0.1830
2.00%	0.9604	0.39775500E-02	0.99584185E-03	0.1811
2.50%	0.9506	0.39371847E-02	0.98571351E-03	0.1793
3.00%	0.9409	0.38970238E-02	0.97563689E-03	0.1774
3.50%	0.9312	0.38570666E-02	0.96561191E-03	0.1756
4.00%	0.9216	0.38173132E-02	0.95563863E-03	0.1738
4.50%	0.9120	0.37777638E-02	0.94571687E-03	0.1720
5.00%	0.9025	0.37384187E-02	0.93584609E-03	0.1702
5.50%	0.8930	0.36992770E-02	0.92602841E-03	0.1684
6.00%	0.8836	0.36603398E-02	0.91626174E-03	0.1666

```
ION OF RETRIEVED RESULTS FOR DATA FIELDS
EX
ND SURNAME
--------------------------------------------------------------------------------------------
                                                                                           !
        SURNAME DECILE  8                                                                  !
-------------------------------------------------------------------------------------------!
                                FILE SIZE                                                  !
-------------------------------------------------------------------------------------------!
      1 MILLION           !           4 MILLION             !        220 MILLION           !
-------------------------------------------------------------------------------------------!
                          !                                 !                              !
    0.41410479E-02        !        0.10368714E-02           !      0.18861800E-04          !
                          !                                 !                              !
    0.40998683E-02        !        0.10265366E-02           !      0.18673657E-04          !
                          !                                 !                              !
    0.40588915E-02        !        0.10162533E-02           !      0.18486456E-04          !
                          !                                 !                              !
    0.40181193E-02        !        0.10060218E-02           !      0.18300198E-04          !
                          !                                 !                              !
    0.39775500E-02        !        0.99584185E-03           !      0.18114885E-04          !
                          !                                 !                              !
    0.39371847E-02        !        0.98571351E-03           !      0.17930511E-04          !
                          !                                 !                              !
    0.38970238E-02        !        0.97563689E-03           !      0.17747083E-04          !
                          !                                 !                              !
    0.38570666E-02        !        0.96561191E-03           !      0.17564597E-04          !
                          !                                 !                              !
    0.38173132E-02        !        0.95563863E-03           !      0.17383053E-04          !
                          !                                 !                              !
    0.37777638E-02        !        0.94571687E-03           !      0.17202456E-04          !
                          !                                 !                              !
    0.37384187E-02        !        0.93584609E-03           !      0.17022799E-04          !
                          !                                 !                              !
    0.36992770E-02        !        0.92602841E-03           !      0.16844086E-04          !
                          !                                 !                              !
    0.36603398E-02        !        0.91626174E-03           !      0.16666315E-04          !
--------------------------------------------------------------------------------------------
```

Table 1i. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
AND SURNAME

SURNAME DECILE 9

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.25503722E-02	0.63820294E-03	0.1160
0.50%	0.9900	0.25249805E-02	0.63183983E-03	0.1149
1.00%	0.9801	0.24997149E-02	0.62550864E-03	0.1137
1.50%	0.9702	0.24745757E-02	0.61920929E-03	0.1126
2.00%	0.9604	0.24495621E-02	0.61294164E-03	0.1114
2.50%	0.9506	0.24246756E-02	0.60670597E-03	0.1103
3.00%	0.9409	0.23999149E-02	0.60050203E-03	0.1092
3.50%	0.9312	0.23752808E-02	0.59432994E-03	0.1080
4.00%	0.9216	0.23507733E-02	0.58810982E-03	0.1069
4.50%	0.9120	0.23263918E-02	0.58208140E-03	0.1058
5.00%	0.9025	0.23021370E-02	0.57600486E-03	0.1047
5.50%	0.8930	0.22780080E-02	0.56996015E-03	0.1036
6.00%	0.8836	0.22540056E-02	0.56394724E-03	0.1025

```
ECISION OF RETRIEVED RESULTS FOR DATA FIELDS
  SEX
  AND SURNAME
-------------------------------------------------------------------------------
          SURNAME DECILE  9
-------------------------------------------------------------------------------
                                FILE SIZE
 !-----------------------------------------------------------------------------
 !     1 MILLION         !          4 MILLION          !      220 MILLION      !
-------------------------------------------------------------------------------
 !   0.25563722E-02      !       0.63820294E-03        !   0.11607321E-04      !
 !   0.25249805E-02      !       0.63183983E-03        !   0.11491540E-04      !
 !   0.24997149E-02      !       0.62550864E-03        !   0.11376339E-04      !
 !   0.24745757E-02      !       0.61920929E-03        !   0.11261716E-04      !
 !   0.24495621E-02      !       0.61294164E-03        !   0.11147676E-04      !
 !   0.24246756E-02      !       0.60670597E-03        !   0.11034215E-04      !
 !   0.23999149E-02      !       0.60050203E-03        !   0.10921335E-04      !
 !   0.23752808E-02      !       0.59432994E-03        !   0.10809034E-04      !
 !   0.23507733E-02      !       0.58818982E-03        !   0.10697315E-04      !
 !   0.23263918E-02      !       0.58208140E-03        !   0.10586176E-04      !
 !   0.23021370E-02      !       0.57600486E-03        !   0.10475617E-04      !
 !   0.22780080E-02      !       0.56996015E-03        !   0.10365639E-04      !
 !   0.22540056E-02      !       0.56394724E-03        !   0.10256239E-04      !
-------------------------------------------------------------------------------
```

Table 1j. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
AND SURNAME

SURNAME DECILE 10				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.59084252E-03	0.14774326E-03	0.26864
0.50%	0.9900	0.58495150E-03	0.14626968E-03	0.26596
1.00%	0.9801	0.57908987E-03	0.14480348E-03	0.26329
1.50%	0.9702	0.57325778E-03	0.14334468E-03	0.26064
2.00%	0.9604	0.56745503E-03	0.14189323E-03	0.2580
2.50%	0.9506	0.56168192E-03	0.14044921E-03	0.2553
3.00%	0.9409	0.55593825E-03	0.13901254E-03	0.2527
3.50%	0.9312	0.55022405E-03	0.13758326E-03	0.2501
4.00%	0.9216	0.54453930E-03	0.13616138E-03	0.2475
4.50%	0.9120	0.53888403E-03	0.13474685E-03	0.2450
5.00%	0.9025	0.53325828E-03	0.13333973E-03	0.2424
5.50%	0.8930	0.52766196E-03	0.13193998E-03	0.2399
6.00%	0.8836	0.52209513E-03	0.13054761E-03	0.2373

SION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
AND SURNAME

SURNAME DECILE 10

FILE SIZE		
1 MILLION	4 MILLION	220 MILLION
0.59084252E-03	0.14774326E-03	0.26864349E-05
0.58495150E-03	0.14626968E-03	0.26596377E-05
0.57908987E-03	0.14480348E-03	0.26329749E-05
0.57325778E-03	0.14334468E-03	0.26064464E-05
0.56745503E-03	0.14189323E-03	0.25800524E-05
0.56168192E-03	0.14044921E-03	0.25537924E-05
0.55593825E-03	0.13901254E-03	0.25276669E-05
0.55022405E-03	0.13758326E-03	0.25016758E-05
0.54453930E-03	0.13616138E-03	0.24758187E-05
0.53888403E-03	0.13474685E-03	0.24500962E-05
0.53325828E-03	0.13333973E-03	0.24245081E-05
0.52766196E-03	0.13193998E-03	0.23990542E-05
0.52209513E-03	0.13054761E-03	0.23737345E-05

Table 2a. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NUMBER
AND SURNAME

		SURNAME DECILE 1		
		FILE SIZE		
ERROR RATE	RECALL FACTOR	1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.99793
0.50%	0.9900	0.10000000E 01	0.10000000E 01	0.99791
1.00%	0.9801	0.10000000E 01	0.10000000E 01	0.99789
1.50%	0.9702	0.10000000E 01	0.10000000E 01	0.99786
2.00%	0.9604	0.10000000E 01	0.10000000E 01	0.99784
2.50%	0.9506	0.10000000E 01	0.10000000E 01	0.99782
3.00%	0.9409	0.10000000E 01	0.10000000E 01	0.99780
3.50%	0.9312	0.10000000E 01	0.10000000E 01	0.99777
4.00%	0.9216	0.10000000E 01	0.10000000E 01	0.99775
4.50%	0.9120	0.10000000E 01	0.10000000E 01	0.99772
5.00%	0.9025	0.10000000E 01	0.10000000E 01	0.99770
5.50%	0.8930	0.10000000E 01	0.10000000E 01	0.99767
6.00%	0.8836	0.10000000E 01	0.10000000E 01	0.99765

```
ECISION OF RETRIEVED RESULTS FOR DATA FIELDS
   STREET NUMBER
   AND SURNAME
-----------------------------------------------------------------------------------
                                                                                  !
        SURNAME DECILE  !                                                         !
----------------------------------------------------------------------------------!
 !                          FILE SIZE                                             !
 !--------------------------------------------------------------------------------!
 !        1 MILLION        !        4 MILLION         !      220 MILLION          !
----------------------------------------------------------------------------------!
 !                         !                          !                           !
 !      0.10000000E 01     !      0.10000000E 01      !      0.99793297E 00       !
 !                         !                          !                           !
 !      0.10000000E 01     !      0.10000000E 01      !      0.99791169E 00       !
 !                         !                          !                           !
 !      0.10000000E 01     !      0.10000000E 01      !      0.99789008E 00       !
 !                         !                          !                           !
 !      0.10000000E 01     !      0.10000000E 01      !      0.99786813E 00       !
 !                         !                          !                           !
 !      0.10000000E 01     !      0.10000000E 01      !      0.99784585E 00       !
 !                         !                          !                           !
 !      0.10000000E 01     !      0.10000000E 01      !      0.99782323E 00       !
 !                         !                          !                           !
 !      0.10000000E 01     !      0.10000000E 01      !      0.99780025E 00       !
 !                         !                          !                           !
 !      0.10000000E 01     !      0.10000000E 01      !      0.99777691E 00       !
 !                         !                          !                           !
 !      0.10000000E 01     !      0.10000000E 01      !      0.99775320E 00       !
 !                         !                          !                           !
 !      0.10000000E 01     !      0.10000000E 01      !      0.99772912E 00       !
 !                         !                          !                           !
 !      0.10000000E 01     !      0.10000000E 01      !      0.99770466E 00       !
 !                         !                          !                           !
 !      0.10000000E 01     !      0.10000000E 01      !      0.99767980E 00       !
 !                         !                          !                           !
 !      0.10000000E 01     !      0.10000000E 01      !      0.99765455E 00       !
-----------------------------------------------------------------------------------
```

Table 2b. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NUMBER
AND SURNAME

SURNAME DECILE 2				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.99985558E 00	0.99911939E 00	0.94877
0.50%	0.9900	0.99985362E 00	0.99911001E 00	0.94828
1.00%	0.9801	0.99985162E 00	0.99910040E 00	0.94778
1.50%	0.9702	0.99984960E 00	0.99909081E 00	0.94727
2.00%	0.9604	0.99984753E 00	0.99908095E 00	0.94676
2.50%	0.9506	0.99984544E 00	0.99907101E 00	0.94625
3.00%	0.9409	0.99984331E 00	0.99906098E 00	0.94572
3.50%	0.9312	0.99984114E 00	0.99905058E 00	0.94519
4.00%	0.9216	0.99983894E 00	0.99904013E 00	0.94464
4.50%	0.9120	0.99983669E 00	0.99902950E 00	0.94410
5.00%	0.9025	0.99983441E 00	0.99901870E 00	0.9435
5.50%	0.8930	0.99983209E 00	0.99900773E 00	0.94297
6.00%	0.8836	0.99982973E 00	0.99899658E 00	0.94240

```
ION OF RETRIEVED RESULTS FOR DATA FIELDS
TREET NUMBER
ND SURNAME
------------------------------------------------------------------------------------
                                                                                   !
      SURNAME DECILE  2                                                            !
-----------------------------------------------------------------------------------!
                              FILE SIZE                                            !
-----------------------------------------------------------------------------------!
      1 MILLION          !            4 MILLION            !      220 MILLION      !
-----------------------------------------------------------------------------------!
                         !                                 !                       !
    0.99985558E 00       !         0.99911939E 00          !   0.94[illegible]7060E 00     !
    0.99985362E 00       !         0.99911001E 00          !   0.94828068E 00      !
    0.99985162E 00       !         0.99910048E 00          !   0.94778381E 00      !
    0.99984960E 00       !         0.99909081E 00          !   0.94727990E 00      !
    0.99984753E 00       !         0.99908095E 00          !   0.94676980E 00      !
    0.99984544E 00       !         0.99907101E 00          !   0.94625038E 00      !
    0.99984331E 00       !         0.99906088E 00          !   0.94572449E 00      !
    0.99984114E 00       !         0.99905058E 00          !   0.94519101E 00      !
    0.99983894E 00       !         0.99904013E 00          !   0.94464979E 00      !
    0.99983669E 00       !         0.99902950E 00          !   0.94410067E 00      !
    0.99983441E 00       !         0.99901878E 00          !   0.94354351E 00      !
    0.99983209E 00       !         0.99900773E 00          !   0.94297814E 00      !
    0.99982973E 00       !         0.99899658E 00          !   0.94240444E 00      !
------------------------------------------------------------------------------------
```

Table 2c. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NUMBER
AND SURNAME

SURNAME DECILE 3

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.99914586E 00	0.99626935E 00	0.82517
0.50%	0.9900	0.99913672E 00	0.99623136E 00	0.82372
1.00%	0.9801	0.99912744E 00	0.99619279E 00	0.82225
1.50%	0.9702	0.99911801E 00	0.99615363E 00	0.82077
2.00%	0.9604	0.99910844E 00	0.99611387E 00	0.81927
2.50%	0.9506	0.99909871E 00	0.99607350E 00	0.81775
3.00%	0.9409	0.99908884E 00	0.99603250E 00	0.81621
3.50%	0.9312	0.99907880E 00	0.99599086E 00	0.81465
4.00%	0.9216	0.99906861E 00	0.99594858E 00	0.81308
4.50%	0.9120	0.99905825E 00	0.99590562E 00	0.81148
5.00%	0.9025	0.99904772E 00	0.99586139E 00	0.80987
5.50%	0.8930	0.99903703E 00	0.99581766E 00	0.80824
6.00%	0.8836	0.99902616E 00	0.99577263E 00	0.80659

ISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NUMBER
AND SURNAME

SURNAME DECILE 3

FILE SIZE		
1 MILLION	4 MILLION	220 MILLION
0.99914586E 00	0.99626935E 00	0.82517753E 00
0.99913672E 00	0.99623136E 00	0.82372623E 00
0.99912744E 00	0.99619279E 00	0.82225809E 00
0.99911801E 00	0.99615363E 00	0.82077287E 00
0.99910844E 00	0.99611387E 00	0.81927034E 00
0.99909871E 00	0.99607350E 00	0.81775026E 00
0.99908884E 00	0.99603250E 00	0.81621237E 00
0.99907880E 00	0.99599086E 00	0.81465642E 00
0.99906861E 00	0.99594858E 00	0.81308216E 00
0.99905825E 00	0.99590562E 00	0.81148930E 00
0.99904772E 00	0.99586199E 00	0.80987762E 00
0.99903763E 00	0.99581766E 00	0.80824683E 00
0.99902616E 00	0.99577263E 00	0.80659664E 00

Table 2d. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NUMBER
AND SURNAME

SURNAME DECILE 4

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.99741254E 00	0.98935990E 00	0.62564
0.50%	0.9900	0.99738591E 00	0.98925322E 00	0.62329
1.00%	0.9801	0.99735088E 00	0.98914494E 00	0.62092
1.50%	0.9702	0.99733142E 00	0.98903503E 00	0.61853
2.00%	0.9604	0.99730355E 00	0.98892346E 00	0.61613
2.50%	0.9506	0.99727524E 00	0.98881018E 00	0.61370
3.00%	0.9409	0.99724650E 00	0.98869518E 00	0.61126
3.50%	0.9312	0.99721730E 00	0.98857840E 00	0.6088
4.00%	0.9216	0.99718764E 00	0.98845982E 00	0.60633
4.50%	0.9120	0.99715752E 00	0.98833940E 00	0.60383
5.00%	0.9025	0.99712691E 00	0.98821710E 00	0.60132
5.50%	0.8930	0.99709582E 00	0.98809288E 00	0.59878
6.00%	0.8836	0.99706423E 00	0.98796670E 00	0.59623

```
RIEVED RESULTS FOR DATA FIELDS
ER

------------------------------------------------------------------------------
RNAME DECILE  4
------------------------------------------------------------------------------
                 FILE SIZE
------------------------------------------------------------------------------
LLION          !          4 MILLION           !        220 MILLION           !
------------------------------------------------------------------------------
               !                              !                              !
41254E 00      !       0.98935990E 00         !      0.62564351E 00          !
38591E 00      !       0.98925322E 00         !      0.62329232E 00          !
35080E 00      !       0.98914494E 00         !      0.62092340E 00          !
33142E 00      !       0.98903503E 00         !      0.61853668E 00          !
30355E 00      !       0.98892346E 00         !      0.61613201E 00          !
27524E 00      !       0.98881018E 00         !      0.61370933E 00          !
24650E 00      !       0.98869518E 00         !      0.61126851E 00          !
21730E 00      !       0.98857840E 00         !      0.60880944E 00          !
18764E 00      !       0.98845982E 00         !      0.60633200E 00          !
15752E 00      !       0.98833940E 00         !      0.60383609E 00          !
12691E 00      !       0.98821710E 00         !      0.60132165E 00          !
09582E 00      !       0.98809280E 00         !      0.59878853E 00          !
06423E 00      !       0.98796670E 00         !      0.59623664E 00          !
------------------------------------------------------------------------------
```

Table 2e. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS STREET NUMBER AND SURNAME

SURNAME DECILE 5

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.99534926E 00	0.98122841E 00	0.485395
0.50%	0.9900	0.99530189E 00	0.98104215E 00	0.48239
1.00%	0.9801	0.99525380E 00	0.98085312E 00	0.480375
1.50%	0.9702	0.99520498E 00	0.98066128E 00	0.47784
2.00%	0.9604	0.99515541E 00	0.98046657E 00	0.47530
2.50%	0.9506	0.99510508E 00	0.98026893E 00	0.47275
3.00%	0.9409	0.99505397E 00	0.98006830E 00	0.470195
3.50%	0.9312	0.99500206E 00	0.97986463E 00	0.46762
4.00%	0.9216	0.99494934E 00	0.97965785E 00	0.46503
4.50%	0.9120	0.99489579E 00	0.97944790E 00	0.46243
5.00%	0.9025	0.99484140E 00	0.97923471E 00	0.45982
5.50%	0.8930	0.99478614E 00	0.97901822E 00	0.45720
6.00%	0.8836	0.99473000E 00	0.97879836E 00	0.4545

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 STREET NUMBER
 AND SURNAME
-------------------------------------------------------------------------------
        SURNAME DECILE  5                                                     !
-------------------------------------------------------------------------------!
                              FILE SIZE                                       !
-------------------------------------------------------------------------------!
       1 MILLION        !          4 MILLION          !      220 MILLION      !
-------------------------------------------------------------------------------!
                        !                             !                       !
    0.99534926E 00      !       0.98122841E 00        !     0.48539514E 00    !
    0.99530189E 00      !       0.98104215E 00        !     0.48289121E 00    !
    0.99525380E 00      !       0.98085312E 00        !     0.48037555E 00    !
    0.99520498E 00      !       0.98066128E 00        !     0.47784815E 00    !
    0.99515541E 00      !       0.98046657E 00        !     0.47530901E 00    !
    0.99510508E 00      !       0.98026893E 00        !     0.47275818E 00    !
    0.99505397E 00      !       0.98006830E 00        !     0.47019571E 00    !
    0.99500206E 00      !       0.97986463E 00        !     0.46762154E 00    !
    0.99494934E 00      !       0.97965785E 00        !     0.46503574E 00    !
    0.99489579E 00      !       0.97944790E 00        !     0.46243833E 00    !
    0.99484140E 00      !       0.97923471E 00        !     0.45982937E 00    !
    0.99478614E 00      !       0.97901822E 00        !     0.45720882E 00    !
    0.99473000E 00      !       0.97879836E 00        !     0.45457678E 00    !
-------------------------------------------------------------------------------
```

Table 2f. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NUMBER
AND SURNAME

SURNAME DECILE 6				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.99357975E 00	0.97433411E 00	0.40687
0.50%	0.9900	0.99351466E 00	0.97408141E 00	0.40445
1.00%	0.9801	0.99344859E 00	0.97382502E 00	0.40203
1.50%	0.9702	0.99338152E 00	0.97356484E 00	0.39959
2.00%	0.9604	0.99331342E 00	0.97330081E 00	0.39715
2.50%	0.9506	0.99324427E 00	0.97303285E 00	0.39471
3.00%	0.9409	0.99317406E 00	0.97276088E 00	0.39225
3.50%	0.9312	0.99310276E 00	0.97248482E 00	0.38979
4.00%	0.9216	0.99303035E 00	0.97220459E 00	0.38732
4.50%	0.9120	0.99295681E 00	0.97192012E 00	0.38485
5.00%	0.9025	0.99288211E 00	0.97163130E 00	0.38236
5.50%	0.8930	0.99280622E 00	0.97133806E 00	0.37988
6.00%	0.8836	0.99272913E 00	0.97104031E 00	0.37738

```
ION OF RETRIEVED RESULTS FOR DATA FIELDS
TREET NUMBER
ND SURNAME
-------------------------------------------------------------------------
                                                                        !
     SURNAME DECILE  6                                                  !
------------------------------------------------------------------------!
                          FILE SIZE                                     !
------------------------------------------------------------------------!
     1 MILLION        !        4 MILLION        !     220 MILLION       !
------------------------------------------------------------------------!
                      !                         !                       !
   0.99357975E 00     !     0.97433411E 00      !    0.40687224E 00     !
   0.99351466E 00     !     0.97408141E 00      !    0.40445505E 00     !
   0.99344059E 00     !     0.97332502E 00      !    0.40203036E 00     !
   0.99338152E 00     !     0.97356484E 00      !    0.39959819E 00     !
   0.99331342E 00     !     0.97330081E 00      !    0.39715864E 00     !
   0.99324427E 00     !     0.97303285E 00      !    0.39471175E 00     !
   0.99317406E 00     !     0.97276088E 00      !    0.39225758E 00     !
   0.99310276E 00     !     0.97248482E 00      !    0.38979622E 00     !
   0.99303035E 00     !     0.97220459E 00      !    0.38732770E 00     !
   0.99295681E 00     !     0.97192012E 00      !    0.38485211E 00     !
   0.99288211E 00     !     0.97163130E 00      !    0.38236954E 00     !
   0.99280622E 00     !     0.97133806E 00      !    0.37988004E 00     !
   0.99272913E 00     !     0.97104031E 00      !    0.37738366E 00     !
-------------------------------------------------------------------------
```

Table 2g. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NUMBER
AND SURNAME

SURNAME DECILE 7

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.98772796E 00	0.95204205E 00	0.2643
0.50%	0.9900	0.98760474E 00	0.95158122E 00	0.2624
1.00%	0.9801	0.98747967E 00	0.95111382E 00	0.2604
1.50%	0.9702	0.98735273E 00	0.95063977E 00	0.2585
2.00%	0.9604	0.98722386E 00	0.95015891E 00	0.2565
2.50%	0.9506	0.98709303E 00	0.94967114E 00	0.2546
3.00%	0.9409	0.98696020E 00	0.94917630E 00	0.2526
3.50%	0.9312	0.98682533E 00	0.94867428E 00	0.2507
4.00%	0.9216	0.98668839E 00	0.94816492E 00	0.2487
4.50%	0.9120	0.98654931E 00	0.94764810E 00	0.2468
5.00%	0.9025	0.98640807E 00	0.94712367E 00	0.2448
5.50%	0.8930	0.98626462E 00	0.94659148E 00	0.2429
6.00%	0.8836	0.98611891E 00	0.94605139E 00	0.2410

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
  STREET NUMBER
  AND SURNAME
-----------------------------------------------------------------------------
        SURNAME DECILE  7                                                    !
-----------------------------------------------------------------------------!
!                             FILE SIZE                                      !
!----------------------------------------------------------------------------!
!      1 MILLION         !        4 MILLION          !      220 MILLION      !
-----------------------------------------------------------------------------!
!                        !                           !                       !
!    0.98772796E 00      !      0.95204205E 00       !    0.26435625E 00     !
!                        !                           !                       !
!    0.98760474E 00      !      0.95158122E 00       !    0.26241119E 00     !
!                        !                           !                       !
!    0.98747967E 00      !      0.95111382E 00       !    0.26046563E 00     !
!                        !                           !                       !
!    0.98735273E 00      !      0.95063977E 00       !    0.25851968E 00     !
!                        !                           !                       !
!    0.98722386E 00      !      0.95015891E 00       !    0.25657339E 00     !
!                        !                           !                       !
!    0.98709303E 00      !      0.94967114E 00       !    0.25462680E 00     !
!                        !                           !                       !
!    0.98696020E 00      !      0.94917630E 00       !    0.25268008E 00     !
!                        !                           !                       !
!    0.98682533E 00      !      0.94867428E 00       !    0.25073323E 00     !
!                        !                           !                       !
!    0.98668839E 00      !      0.94816492E 00       !    0.24878637E 00     !
!                        !                           !                       !
!    0.98654931E 00      !      0.94764810E 00       !    0.24683952E 00     !
!                        !                           !                       !
!    0.98640807E 00      !      0.94712367E 00       !    0.24489284E 00     !
!                        !                           !                       !
!    0.98626462E 00      !      0.94659148E 00       !    0.24294635E 00     !
!                        !                           !                       !
!    0.98611891E 00      !      0.94605139E 00       !    0.24100017E 00     !
-----------------------------------------------------------------------------
```

Table 2h. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NUMBER
AND SURNAME

SURNAME DECILE 8

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.95454816E 00	0.83878098E 00	0.861863
0.50%	0.9900	0.95410856E 00	0.83741865E 00	0.85400
1.00%	0.9801	0.95366268E 00	0.83604012E 00	0.846163
1.50%	0.9702	0.95321041E 00	0.83464518E 00	0.838352
2.00%	0.9604	0.95275162E 00	0.83323359E 00	0.830567
2.50%	0.9506	0.95228619E 00	0.83180508E 00	0.822609
3.00%	0.9409	0.95181400E 00	0.83035942E 00	0.815070
3.50%	0.9312	0.95133492E 00	0.82889636E 00	0.807373
4.00%	0.9216	0.95084880E 00	0.82741563E 00	0.799695
4.50%	0.9120	0.95035553E 00	0.82591698E 00	0.792045
5.00%	0.9025	0.94985495E 00	0.82440014E 00	0.78442
5.50%	0.8930	0.94934693E 00	0.82286483E 00	0.776826
6.00%	0.8836	0.94883133E 00	0.82131081E 00	0.769257

N OF RETRIEVED RESULTS FOR DATA FIELDS
EET NUMBER
SURNAME

SURNAME DECILE 8		
FILE SIZE		
1 MILLION	4 MILLION	220 MILLION
0.95454816E 00	0.83878098E 00	0.86186388E-01
0.95410856E 00	0.83741865E 00	0.85400079E-01
0.95366268E 00	0.83604012E 00	0.84616359E-01
0.95321041E 00	0.83464518E 00	0.83835265E-01
0.95275162E 00	0.83323359E 00	0.83056799E-01
0.95228619E 00	0.83190508E 00	0.82280972E-01
0.95181409E 00	0.83035942E 00	0.81507821E-01
0.95133492E 00	0.82889676E 00	0.80737344E-01
0.95084880E 00	0.82741563E 00	0.79969570E-01
0.95035553E 00	0.82591698E 00	0.79204514E-01
0.94985495E 00	0.82440014E 00	0.78442185E-01
0.94934693E 00	0.82286483E 00	0.77682605E-01
0.94883133E 00	0.82131081E 00	0.76925794E-01

Table 2i. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NUMBER
AND SURNAME

SURNAME DECILE 9				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.92812776E 00	0.76196447E 00	0.54856
0.50%	0.9900	0.92745231E 00	0.76013881E 00	0.54338
1.00%	0.9801	0.92676760E 00	0.75829436E 00	0.53823
1.50%	0.9702	0.92607344E 00	0.75643086E 00	0.53309
2.00%	0.9604	0.92536966E 00	0.75454810E 00	0.52798
2.50%	0.9506	0.92465612E 00	0.75264586E 00	0.52289
3.00%	0.9409	0.92393261E 00	0.75072389E 00	0.5178
3.50%	0.9312	0.92319897E 00	0.74878199E 00	0.51276
4.00%	0.9216	0.92245500E 00	0.74681992E 00	0.50773
4.50%	0.9120	0.92170053E 00	0.74483743E 00	0.50272
5.00%	0.9025	0.92093335E 00	0.74283427E 00	0.49773
5.50%	0.8930	0.92015927E 00	0.74081024E 00	0.4927
6.00%	0.8836	0.91937208E 00	0.73876505E 00	0.4878

```
N OF RETRIEVED RESULTS FOR DATA FIELDS
EET NUMBER
 SURNAME
--------------------------------------------------------------------------------------
        SURNAME DECILE  9                                                             !
--------------------------------------------------------------------------------------!
                               FILE SIZE                                              !
--------------------------------------------------------------------------------------!
                         !           4 MILLION          !         220 MILLION         !
       1 MILLION         !                              !                             !
--------------------------------------------------------------------------------------!
                         !                              !     0.54856074E-01          !
                         !        0.76196447E 00        !                             !
   0.92612776E 00        !                              !     0.54338607E-01          !
                         !        0.76013881E 00        !                             !
   0.92745231E 00        !                              !     0.53823169E-01          !
                         !        0.75829436E 00        !                             !
   0.92676760E 00        !                              !     0.53309769E-01          !
                         !        0.75643086E 00        !                             !
   0.92607344E 00        !                              !     0.52798416E-01          !
                         !        0.75454810E 00        !                             !
   0.92536966E 00        !                              !     0.52289118E-01          !
                         !        0.75264586E 00        !                             !
   0.92465612E 00        !                              !     0.51781884E-01          !
                         !        0.75072389E 00        !                             !
   0.92393261E 00        !                              !     0.51276712E-01          !
                         !        0.74878199E 00        !                             !
   0.92319897E 00        !                              !     0.50773626E-01          !
                         !        0.74681992E 00        !                             !
   0.92245500E 00        !                              !     0.50272615E-01          !
                         !        0.74483743E 00        !                             !
   0.92170053E 00        !                              !     0.49773703E-01          !
                         !        0.74283427E 00        !                             !
   0.92093535E 00        !                              !     0.49276885E-01          !
                         !        0.74081024E 00        !                             !
   0.92015927E 00        !                              !     0.48782173E-01          !
                         !        0.73876505E 00        !-----------------------------
   0.91937208E 00        !                              
--------------------------------------------------------
```

Table 2j. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NUMBER
AND SURNAME

SURNAME DECILE 10

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MILL
0.00%	1.0000	0.74911006E 00	0.42551081E 00	0.1325477
0.50%	0.9900	0.74721138E 00	0.42305085E 00	0.1312420
1.00%	0.9801	0.74529365E 00	0.42059837E 00	0.1299442
1.50%	0.9702	0.74335661E 00	0.41812942E 00	0.1286517
2.00%	0.9604	0.74140007E 00	0.41565207E 00	0.1273655
2.50%	0.9506	0.73942304E 00	0.41316638E 00	0.1260855
3.00%	0.9409	0.73742767E 00	0.41067232E 00	0.1248117
3.50%	0.9312	0.73541133E 00	0.40817005E 00	0.1235441
4.00%	0.9216	0.73337459E 00	0.40565954E 00	0.1222820
4.50%	0.9120	0.73131725E 00	0.40314092E 00	0.1210277
5.00%	0.9025	0.72923908E 00	0.40061423E 00	0.1197708
5.50%	0.8930	0.72713983E 00	0.39807954E 00	0.1185362
6.00%	0.8836	0.72501929E 00	0.39553692E 00	0.1172998

```
 OF RETRIEVED RESULTS FOR DATA FIELDS
ET NUMBER
 SURNAME
---------------------------------------------------------------------------------
                                                                                !
     SURNAME DECILE 10                                                          !
--------------------------------------------------------------------------------!
                            FILE SIZE                                           !
--------------------------------------------------------------------------------!
   1 MILLION          !          4 MILLION            !      220 MILLION        !
--------------------------------------------------------------------------------!
                      !                               !                         !
 0.74911006E 00       !       0.42551081E 00          !    0.13254772E-01       !
 0.74721138E 00       !       0.42305085E 00          !    0.13124207E-01       !
 0.74529365E 00       !       0.42059837E 00          !    0.12994422E-01       !
 0.74335661E 00       !       0.41812942E 00          !    0.12865179E-01       !
 0.74140007E 00       !       0.41565207E 00          !    0.12736556E-01       !
 0.73942304E 00       !       0.41316638E 00          !    0.12608554E-01       !
 0.73742767E 00       !       0.41067232E 00          !    0.12481174E-01       !
 0.73541133E 00       !       0.40817005E 00          !    0.12354417E-01       !
 0.73337459E 00       !       0.40565954E 00          !    0.12228203E-01       !
 0.73131725E 00       !       0.40314092E 00          !    0.12102771E-01       !
 0.72923908E 00       !       0.40061423E 00          !    0.11977884E-01       !
 0.72713983E 00       !       0.39807954E 00          !    0.11853620E-01       !
 0.72501929E 00       !       0.39553692E 00          !    0.11729982E-01       !
---------------------------------------------------------------------------------
```

Table 3a. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 1

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.95269
0.50%	0.9900	0.10000000E 01	0.10000000E 01	0.95222
1.00%	0.9801	0.10000000E 01	0.10000000E 01	0.95175
1.50%	0.9702	0.10000000E 01	0.10000000E 01	0.95127
2.00%	0.9604	0.10000000E 01	0.10000000E 01	0.95079
2.50%	0.9506	0.10000000E 01	0.10000000E 01	0.95030
3.00%	0.9409	0.10000000E 01	0.10000000E 01	0.94980
3.50%	0.9312	0.10000000E 01	0.10000000E 01	0.94929
4.00%	0.9216	0.10000000E 01	0.10000000E 01	0.94878
4.50%	0.9120	0.10000000E 01	0.10000000E 01	0.94825
5.00%	0.9025	0.10000000E 01	0.10000000E 01	0.94772
5.50%	0.8930	0.10000000E 01	0.10000000E 01	0.94719
6.00%	0.8836	0.10000000E 01	0.10000000E 01	0.94664

```
SION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME
------------------------------------------------------------------------
      SURNAME DECILE  1                                                !
-----------------------------------------------------------------------!
                            FILE SIZE                                  !
-----------------------------------------------------------------------!
   1 MILLION          !       4 MILLION          !     220 MILLION     !
-----------------------------------------------------------------------!
                      !                          !                     !
   0.10000000E 01     !     0.10000000E 01       !   0.95269238E 00    !
                      !                          !                     !
   0.10000000E 01     !     0.10000000E 01       !   0.95222752E 00    !
                      !                          !                     !
   0.10000000E 01     !     0.10000000E 01       !   0.95175601E 00    !
                      !                          !                     !
   0.10000000E 01     !     0.10000000E 01       !   0.95127772E 00    !
                      !                          !                     !
   0.10000000E 01     !     0.10000000E 01       !   0.95079252E 00    !
                      !                          !                     !
   0.10000000E 01     !     0.10000000E 01       !   0.95030029E 00    !
                      !                          !                     !
   0.10000000E 01     !     0.10000000E 01       !   0.94980088E 00    !
                      !                          !                     !
   0.10000000E 01     !     0.10000000E 01       !   0.94929418E 00    !
                      !                          !                     !
   0.10000000E 01     !     0.10000000E 01       !   0.94878003E 00    !
                      !                          !                     !
   0.10000000E 01     !     0.10000000E 01       !   0.94825829E 00    !
                      !                          !                     !
   0.10000000E 01     !     0.10000000E 01       !   0.94772882E 00    !
                      !                          !                     !
   0.10000000E 01     !     0.10000000E 01       !   0.94719146E 00    !
                      !                          !                     !
   0.10000000E 01     !     0.10000000E 01       !   0.94664607E 00    !
------------------------------------------------------------------------
```

Table 3b. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 2

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.99651603E 00	0.97925930E 00	0.43583
0.50%	0.9900	0.99646903E 00	0.97904309E 00	0.43336
1.00%	0.9801	0.99642126E 00	0.97882364E 00	0.43089
1.50%	0.9702	0.99637278E 00	0.97860087E 00	0.42840
2.00%	0.9604	0.99632333E 00	0.97837473E 00	0.42591
2.50%	0.9506	0.99627315E 00	0.97814515E 00	0.42341
3.00%	0.9409	0.99622213E 00	0.97791205E 00	0.42090
3.50%	0.9312	0.99617025E 00	0.97767538E 00	0.41838
4.00%	0.9216	0.99611759E 00	0.97743505E 00	0.41585
4.50%	0.9120	0.99606386E 00	0.97719098E 00	0.41331
5.00%	0.9025	0.99600930E 00	0.97694312E 00	0.41077
5.50%	0.8930	0.99595382E 00	0.97669137E 00	0.40821
6.00%	0.8836	0.99589736E 00	0.97643566E 00	0.40565

```
ON OF RETRIEVED RESULTS FOR DATA FIELDS
TE OF BIRTH (MONTH AND DAY ONLY)
) SURNAME
-------------------------------------------------------------------------------------------
      SUPNAME DECILE  2                                                                   !
-------------------------------------------------------------------------------------------
                              FILE SIZE                                                   !
-------------------------------------------------------------------------------------------
    1 MILLION          !           4 MILLION            !          220 MILLION            !
-------------------------------------------------------------------------------------------
                       !                                !                                 !
  0.99651603E 00       !        0.97925930E 00          !        0.43583044E 00           !
  0.99546903E 00       !        0.97904309E 00          !        0.43336482E 00           !
  0.99642126E 00       !        0.97882364E 00          !        0.43089067E 00           !
  0.99637270E 00       !        0.97860087E 00          !        0.42840628E 00           !
  0.99632333E 00       !        0.97837473E 00          !        0.42591347E 00           !
  0.99627315E 00       !        0.97814515E 00          !        0.42341175E 00           !
  0.99622213E 00       !        0.97791205E 00          !        0.42099106E 00           !
  0.99617025E 00       !        0.97767538E 00          !        0.41838159E 00           !
  0.99611759E 00       !        0.97743505E 00          !        0.41585330E 00           !
  0.99606386E 00       !        0.97719098E 00          !        0.41331630E 00           !
  0.99600930E 00       !        0.97694312E 00          !        0.41077060E 00           !
  0.99595382E 00       !        0.97669137E 00          !        0.40821629E 00           !
  0.99589736E 00       !        0.97643566E 00          !        0.40565345E 00           !
-------------------------------------------------------------------------------------------
```

Table 3c. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 3				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00x	1.0000	0.97972774E 00	0.91744404E 00	0.16449
0.50x	0.9900	0.97951602E 00	0.91667214E 00	0.16312
1.00x	0.9801	0.97930111E 00	0.91588899E 00	0.16175
1.50x	0.9702	0.97908296E 00	0.91509520E 00	0.16038
2.00x	0.9604	0.97886150E 00	0.91429050E 00	0.15901
2.50x	0.9506	0.97863667E 00	0.91347494E 00	0.15765
3.00x	0.9409	0.97840839E 00	0.91264809E 00	0.15629
3.50x	0.9312	0.97817660E 00	0.91180983E 00	0.15493
4.00x	0.9216	0.97794123E 00	0.91095996E 00	0.15358
4.50x	0.9120	0.97770219E 00	0.91009825E 00	0.15222
5.00x	0.9025	0.97745944E 00	0.90922454E 00	0.15087
5.50x	0.8930	0.97721287E 00	0.90833856E 00	0.14952
6.00x	0.8836	0.97696242E 00	0.90744612E 00	0.14818

```
SION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME
-------------------------------------------------------------------------------------
                                                                                    !
         SURNAME DECILE  3                                                          !
-------------------------------------------------------------------------------------
                               FILE SIZE                                            !
-------------------------------------------------------------------------------------
       1 MILLION          !           4 MILLION          !         220 MILLION      !
-------------------------------------------------------------------------------------
                          !                              !                          !
     0.97972774E 00       !       0.91744404E 00         !      0.16449893E 00      !
                          !                              !                          !
     0.97951602E 00       !       0.91667214E 00         !      0.16312542E 00      !
                          !                              !                          !
     0.97930111E 00       !       0.91588899E 00         !      0.16175429E 00      !
                          !                              !                          !
     0.97908296E 00       !       0.91509520E 00         !      0.16038560E 00      !
                          !                              !                          !
     0.97886150E 00       !       0.91429050E 00         !      0.15901940E 00      !
                          !                              !                          !
     0.97863667E 00       !       0.91347494E 00         !      0.15765576E 00      !
                          !                              !                          !
     0.97840839E 00       !       0.91264809E 00         !      0.15629467E 00      !
                          !                              !                          !
     0.97817660E 00       !       0.91180983E 00         !      0.15493624E 00      !
                          !                              !                          !
     0.97794123E 00       !       0.91095996E 00         !      0.15358048E 00      !
                          !                              !                          !
     0.97770219E 00       !       0.91009825E 00         !      0.15222746E 00      !
                          !                              !                          !
     0.97745944E 00       !       0.90922454E 00         !      0.15087719E 00      !
                          !                              !                          !
     0.97721287E 00       !       0.90833856E 00         !      0.14952976E 00      !
                          !                              !                          !
     0.97696242E 00       !       0.90744012E 00         !      0.14818517E 00      !
-------------------------------------------------------------------------------------
```

Table 3d. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS DATE OF BIRTH (MONTH AND DAY ONLY) AND SURNAME

SURNAME DECILE 4

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.94091883E 00	0.79463892E 00	0.65168
0.50%	0.9900	0.94034824E 00	0.79299041E 00	0.64560
1.00%	0.9801	0.93976964E 00	0.79132381E 00	0.63954
1.50%	0.9702	0.93918287E 00	0.78963800E 00	0.63351
2.00%	0.9604	0.93858780E 00	0.78793538E 00	0.62749
2.50%	0.9506	0.93798426E 00	0.78621336E 00	0.62150
3.00%	0.9409	0.93737210E 00	0.78447170E 00	0.61554
3.50%	0.9312	0.93675116E 00	0.78271104E 00	0.60959
4.00%	0.9216	0.93612127E 00	0.78093082E 00	0.60367
4.50%	0.9120	0.93548227E 00	0.77913079E 00	0.5977
5.00%	0.9025	0.93483399E 00	0.77731071E 00	0.5919
5.50%	0.8930	0.93417624E 00	0.77547030E 00	0.5860
6.00%	0.8836	0.93350886E 00	0.77360931E 00	0.5802

```
ON OF RETRIEVED RESULTS FOR DATA FIELDS
TE OF BIRTH (MONTH AND DAY ONLY)
 SURNAME
---------------------------------------------------------------------------------
                                                                                !
       SURNAME DECILE  4                                                        !
--------------------------------------------------------------------------------!
                              FILE SIZE                                         !
--------------------------------------------------------------------------------!
      1 MILLION        !           4 MILLION            !      220 MILLION      !
--------------------------------------------------------------------------------!
                       !                                !                       !
  0.94091883E 00       !       0.79463892E 00           !    0.65168938E-01     !
                       !                                !                       !
  0.94034824E 00       !       0.79299041E 00           !    0.64560803E-01     !
                       !                                !                       !
  0.93976964E 00       !       0.79132391E 00           !    0.63954923E-01     !
                       !                                !                       !
  0.93918287E 00       !       0.78963808E 00           !    0.63351318E-01     !
                       !                                !                       !
  0.93858780E 00       !       0.78793538E 00           !    0.62749990E-01     !
                       !                                !                       !
  0.93798426E 00       !       0.78621336E 00           !    0.62150970E-01     !
                       !                                !                       !
  0.93737210E 00       !       0.78447170E 00           !    0.61554240E-01     !
                       !                                !                       !
  0.93675116E 00       !       0.78271104E 00           !    0.60959835E-01     !
                       !                                !                       !
  0.93612127E 00       !       0.78093082E 00           !    0.60367748E-01     !
                       !                                !                       !
  0.93548227E 00       !       0.77913079E 00           !    0.59778004E-01     !
                       !                                !                       !
  0.93483099E 00       !       0.77731071E 00           !    0.59190601E-01     !
                       !                                !                       !
  0.93417624E 00       !       0.77547030E 00           !    0.58605558E-01     !
                       !                                !                       !
  0.93350886E 00       !       0.77360931E 00           !    0.58022877E-01     !
---------------------------------------------------------------------------------
```

Table 3e. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS DATE OF BIRTH (MONTH AND DAY ONLY) AND SURNAME

SURNAME DECILE 5

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.89839557E 00	0.68506567E 00	0.37855
0.50%	0.9900	0.89746686E 00	0.68289299E 00	0.37491
1.00%	0.9801	0.89652596E 00	0.68070133E 00	0.37130
1.50%	0.9702	0.89557267E 00	0.67849053E 00	0.36769
2.00%	0.9604	0.89460677E 00	0.67626042E 00	0.36410
2.50%	0.9506	0.89362805E 00	0.67401084E 00	0.36053
3.00%	0.9409	0.89263628E 00	0.67174161E 00	0.35697
3.50%	0.9312	0.89163126E 00	0.66945257E 00	0.35343
4.00%	0.9216	0.89061276E 00	0.66714355E 00	0.34991
4.50%	0.9120	0.88958054E 00	0.66481438E 00	0.34640
5.00%	0.9025	0.88853437E 00	0.66246494E 00	0.34290
5.50%	0.8930	0.88747401E 00	0.66009498E 00	0.33943
6.00%	0.8836	0.88639921E 00	0.65770440E 00	0.3359

```
SION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME
-------------------------------------------------------------------------------------
        SURNAME DECILE  5                                                           !
------------------------------------------------------------------------------------!
                              FILE SIZE                                             !
------------------------------------------------------------------------------------!
      1 MILLION          !           4 MILLION          !       220 MILLION         !
------------------------------------------------------------------------------------!
                         !                              !                           !
    0.89839557E 00       !       0.68506567E 00         !     0.37855398E-01        !
    0.89746686E 00       !       0.68289299E 00         !     0.37491931E-01        !
    0.89652596E 00       !       0.68070133E 00         !     0.37130016E-01        !
    0.89557267E 00       !       0.67849053E 00         !     0.36769651E-01        !
    0.89460677E 00       !       0.67626042E 00         !     0.36410838E-01        !
    0.89362805E 00       !       0.67401084E 00         !     0.36053595E-01        !
    0.89263628E 00       !       0.67174161E 00         !     0.35697911E-01        !
    0.89163126E 00       !       0.66945257E 00         !     0.35343797E-01        !
    0.89061276E 00       !       0.66714355E 00         !     0.34991253E-01        !
    0.88950054E 00       !       0.66481438E 00         !     0.34640291E-01        !
    0.88853437E 00       !       0.66246494E 00         !     0.34290907E-01        !
    0.88747401E 00       !       0.66009498E 00         !     0.33943105E-01        !
    0.88639921E 00       !       0.65770440E 00         !     0.33596890E-01        !
-------------------------------------------------------------------------------------
```

Table 3f. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS DATE OF BIRTH (MONTH AND DAY ONLY) AND SURNAME

SURNAME DECILE 6				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00x	1.0000	0.86474967E 00	0.61236893E 00	0.27817
0.50x	0.9900	0.86356358E 00	0.60998191E 00	0.27547
1.00x	0.9801	0.86236277E 00	0.60757764E 00	0.27279
1.59x	0.9702	0.86114700E 00	0.60515590E 00	0.27011
2.00x	0.9604	0.85991602E 00	0.60271672E 00	0.26745
2.50x	0.9506	0.85866958E 00	0.60025985E 00	0.26480
3.00x	0.9409	0.85740748E 00	0.59778531E 00	0.26216
3.50x	0.9312	0.85612944E 00	0.59529298E 00	0.25954
4.00x	0.9216	0.85483524E 00	0.59279278E 00	0.25692
4.50x	0.9120	0.85352460E 00	0.59025460E 00	0.25432
5.00x	0.9025	0.85219728E 00	0.58770836E 00	0.25173
5.50x	0.8930	0.85085299E 00	0.58514395E 00	0.24916
6.00x	0.8836	0.84949146E 00	0.58256132E 00	0.24659

le 3f. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 6			
ALL FACTOR	FILE SIZE		
	1 MILLION	4 MILLION	220 MILLION
1.0000	0.86474967E 00	0.61236893E 00	0.27817838E-01
0.9900	0.86356358E 00	0.60998191E 00	0.27547992E-01
0.9801	0.86236277E 00	0.60757764E 00	0.27279351E-01
0.9702	0.86114700E 00	0.60515590E 00	0.27011910E-01
0.9604	0.85991602E 00	0.60271672E 00	0.26745680E-01
0.9506	0.85866958E 00	0.60025985E 00	0.26480663E-01
0.9409	0.85740748E 00	0.59778531E 00	0.26210852E-01
0.9312	0.85612944E 00	0.59529298E 00	0.25954263E-01
0.9216	0.85483524E 00	0.59278278E 00	0.25692885E-01
0.9120	0.85352460E 00	0.59025460E 00	0.25432731E-01
0.9025	0.85219728E 00	0.58770836E 00	0.25173795E-01
0.8930	0.85085299E 00	0.58514395E 00	0.24916084E-01
0.8836	0.84949146E 00	0.58256132E 00	0.24659595E-01

Table 3g. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

		SURNAME DECILE 7		
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220
0.00%	1.0000	0.76879851E 00	0.45238829E 00	0.147
0.50%	0.9900	0.76700429E 00	0.44990337E 00	0.146
1.00%	0.9801	0.76519129E 00	0.44740843E 00	0.144
1.50%	0.9702	0.76335937E 00	0.44490360E 00	0.143
2.00%	0.9604	0.76150824E 00	0.44238883E 00	0.141
2.50%	0.9506	0.75963768E 00	0.43986419E 00	0.140
3.00%	0.9409	0.75774745E 00	0.43732968E 00	0.139
3.50%	0.9312	0.75583734E 00	0.43470541E 00	0.137
4.00%	0.9216	0.75390710E 00	0.43223137E 00	0.136
4.50%	0.9120	0.75195647E 00	0.42966765E 00	0.134
5.00%	0.9025	0.74998524E 00	0.42709431E 00	0.133
5.50%	0.8930	0.74799316E 00	0.42451133E 00	0.132
6.00%	0.8836	0.74597997E 00	0.42191085E 00	0.130

```
ION OF RETRIEVED RESULTS FOR DATA FIELDS
ATE OF BIRTH (MONTH AND DAY ONLY)
ND SURNAME
------------------------------------------------------------------------------------
        SURNAME DECILE  7                                                          !
------------------------------------------------------------------------------------
                             FILE SIZE                                             !
------------------------------------------------------------------------------------
     1 MILLION        !           4 MILLION          !       220 MILLION           !
------------------------------------------------------------------------------------
   0.76879851E 00     !        0.45238829E 00        !     0.14768161E-01          !
   0.76700429E 00     !        0.44990337E 00        !     0.14623001E-01          !
   0.76519129E 00     !        0.44740843E 00        !     0.14478525E-01          !
   0.76335937E 00     !        0.44490360E 00        !     0.14334736E-01          !
   0.76150824E 00     !        0.44238883E 00        !     0.14191632E-01          !
   0.75963768E 00     !        0.43986419E 00        !     0.14049215E-01          !
   0.75774745E 00     !        0.43732968E 00        !     0.13907485E-01          !
   0.75583734E 00     !        0.43478541E 00        !     0.13766445E-01          !
   0.75390710E 00     !        0.43223139E 00        !     0.13626094E-01          !
   0.75195647E 00     !        0.42966765E 00        !     0.13486433E-01          !
   0.74998524E 00     !        0.42709431E 00        !     0.13347461E-01          !
   0.74799316E 00     !        0.42451133E 00        !     0.13209180E-01          !
   0.74597997E 00     !        0.42191885E 00        !     0.13071591E-01          !
------------------------------------------------------------------------------------
```

Table 3h. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 8

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.46456954E 00	0.17797547E 00	0.391869
0.50%	0.9900	0.46207365E 00	0.17651307E 00	0.387976
1.00%	0.9901	0.45956712E 00	0.17505284E 00	0.384101
1.50%	0.9702	0.45704998E 00	0.17359483E 00	0.380246
2.00%	0.9604	0.45452228E 00	0.17213909E 00	0.376410
2.50%	0.9506	0.45198396E 00	0.17068564E 00	0.372593
3.00%	0.9409	0.44943519E 00	0.16923459E 00	0.368795
3.50%	0.9312	0.44687590E 00	0.16778595E 00	0.365017
4.00%	0.9216	0.44430618E 00	0.16633978E 00	0.361258
4.50%	0.9120	0.44172693E 00	0.16489612E 00	0.357518
5.00%	0.9025	0.43913558E 00	0.16345507E 00	0.353797
5.50%	0.8930	0.43653474E 00	0.16201661E 00	0.350096
6.00%	0.8836	0.43392369E 00	0.16058086E 00	0.346414

SION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 8

FILE SIZE		
1 MILLION	4 MILLION	220 MILLION
0.46456954E 00	0.17797547E 00	0.39186970E-02
0.46207365E 00	0.17651307E 00	0.38797607E-02
0.45956712E 00	0.17505284E 00	0.38410151E-02
0.45704998E 00	0.17359483E 00	0.38024621E-02
0.45452228E 00	0.17213909E 00	0.37641011E-02
0.45198396E 00	0.17068564E 00	0.37259325E-02
0.44943519E 00	0.16923459E 00	0.36879557E-02
0.44687590E 00	0.16778595E 00	0.36501718E-02
0.44430618E 00	0.16633978E 00	0.36125801E-02
0.44172693E 00	0.16489612E 00	0.35751818E-02
0.43913558E 00	0.16345507E 00	0.35379750E-02
0.43653474E 00	0.16201661E 00	0.35009613E-02
0.43392369E 00	0.16058086E 00	0.34641401E-02

Table 3i PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS DATE OF BIRTH (MONTH AND DAY ONLY) AND SURNAME

SURNAME DECILE 9

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.34790396E 00	0.11755083E 00	0.2415
0.50%	0.9900	0.34563116E 00	0.11651465E 00	0.2391
1.00%	0.9801	0.34335402E 00	0.11548128E 00	0.2367
1.50%	0.9702	0.34107258E 00	0.11445070E 00	0.2343
2.00%	0.9604	0.33878694E 00	0.11342295E 00	0.2319
2.50%	0.9506	0.33649717E 00	0.11239807E 00	0.2296
3.00%	0.9409	0.33420334E 00	0.11137609E 00	0.2272
3.50%	0.9312	0.33190555E 00	0.11035701E 00	0.2249
4.00%	0.9216	0.32960390E 00	0.10934090E 00	0.2226
4.50%	0.9120	0.32729841E 00	0.10832775E 00	0.2203
5.00%	0.9025	0.32498924E 00	0.10731762E 00	0.2180
5.50%	0.8930	0.32267639E 00	0.10631051E 00	0.2157
6.00%	0.8836	0.32036002E 00	0.10530646E 00	0.2134

ON OF RETRIEVED RESULTS FOR DATA FIELDS
TE OF BIRTH (MONTH AND DAY ONLY)
D SURNAME

SURNAME DECILE 9

FILE SIZE		
1 MILLION	4 MILLION	220 MILLION
0.34790396E 00	0.11755083E 00	0.24151414E-02
0.34563116E 00	0.11651465E 00	0.23911081E-02
0.34335402E 00	0.11548128E 00	0.23671940E-02
0.34107258E 00	0.11445070E 00	0.23433990E-02
0.33878694E 00	0.11342295E 00	0.23197234E-02
0.33649717E 00	0.11239807E 00	0.22961675E-02
0.33420334E 00	0.11137609E 00	0.22727305E-02
0.33190555E 00	0.11035701E 00	0.22494131E-02
0.32960390E 00	0.10934090E 00	0.22262153E-02
0.32729841E 00	0.10932775E 00	0.22031370E-02
0.32498924E 00	0.10731762E 00	0.21801780E-02
0.32267639E 00	0.10631051E 00	0.21573385E-02
0.32036002E 00	0.10530646E 00	0.21346184E-02

Table 3j. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 10				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.10981100E 00	0.29901143E-01	0.5690
0.50%	0.9900	0.10883451E 00	0.29611683E-01	0.5544
1.00%	0.9801	0.10786078E 00	0.29323506E-01	0.5488
1.50%	0.9702	0.10688985E 00	0.29036610E-01	0.5433
2.00%	0.9604	0.10592175E 00	0.28751000E-01	0.5378
2.50%	0.9506	0.10495649E 00	0.28466679E-01	0.5323
3.00%	0.9409	0.10399411E 00	0.28183645E-01	0.5269
3.50%	0.9312	0.10393462E 00	0.27901903E-01	0.5215
4.00%	0.9216	0.10207805E 00	0.27621455E-01	0.5161
4.50%	0.9120	0.10112442E 00	0.27342301E-01	0.5107
5.00%	0.9025	0.10017378E 00	0.27064452E-01	0.5054
5.50%	0.8930	0.99226130E-01	0.26787900E-01	0.5001
6.00%	0.8836	0.98281509E-01	0.26512654E-01	0.4948

```
ION OF RETRIEVED RESULTS FOR DATA FIELDS
ATE OF BIRTH (MONTH AND DAY ONLY)
ND SURNAME
--------------------------------------------------------------------------------
        SURNAME DECILE 10                                                      !
-------------------------------------------------------------------------------!
                              FILE SIZE                                        !
-------------------------------------------------------------------------------!
      1 MILLION        !         4 MILLION          !        220 MILLION       !
-------------------------------------------------------------------------------!
                       !                            !                          !
    0.10981100E 00     !      0.29901143E-01        !     0.56000237E-03       !
    0.10883451E 00     !      0.29611683E-01        !     0.55441947E-03       !
    0.10786078E 00     !      0.29323506E-01        !     0.54886444E-03       !
    0.10688985E 00     !      0.29036610E-01        !     0.54333740E-03       !
    0.10592175E 00     !      0.28751000E-01        !     0.53783822E-03       !
    0.10495649E 00     !      0.28466679E-01        !     0.53236699E-03       !
    0.10399411E 00     !      0.28183645E-01        !     0.52692361E-03       !
    0.10303462E 00     !      0.27901903E-01        !     0.52150825E-03       !
    0.10207805E 00     !      0.27621455E-01        !     0.51612079E-03       !
    0.10112442E 00     !      0.27342301E-01        !     0.51076132E-03       !
    0.10017378E 00     !      0.27064452E-01        !     0.50542974E-03       !
    0.99226130E-01     !      0.26787900E-01        !     0.50012608E-03       !
    0.98281509E-01     !      0.26512654E-01        !     0.49485038E-03       !
--------------------------------------------------------------------------------
```

Table 4a. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS DATE OF BIRTH (YEAR, MONTH, AND DAY) AND SURNAME

SURNAME DECILE 1

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.998
0.50%	0.9900	0.10000000E 01	0.10000000E 01	0.998
1.00%	0.9801	0.10000000E 01	0.10000000E 01	0.998
1.50%	0.9702	0.10000000E 01	0.10000000E 01	0.998
2.00%	0.9604	0.10000000E 01	0.10000000E 01	0.998
2.50%	0.9506	0.10000000E 01	0.10000000E 01	0.998
3.00%	0.9409	0.10000000E 01	0.10000000E 01	0.998
3.50%	0.9312	0.10000000E 01	0.10000000E 01	0.998
4.00%	0.9216	0.10000000E 01	0.10000000E 01	0.998
4.50%	0.9120	0.10000000E 01	0.10000000E 01	0.998
5.00%	0.9025	0.10000000E 01	0.10000000E 01	0.998
5.50%	0.8930	0.10000000E 01	0.10000000E 01	0.998
6.00%	0.8836	0.10000000E 01	0.10000000E 01	0.998

```
CISION OF RETRIEVED RESULTS FOR DATA FIELDS
 DATE OF BIRTH (YEAR, MONTH, AND DAY)
 AND SURNAME
----------------------------------------------------------------------------------
                                                                                  !
        SURNAME DECILE  1                                                         !
----------------------------------------------------------------------------------!
!                              FILE SIZE                                          !
!---------------------------------------------------------------------------------!
!       1 MILLION          !         4 MILLION          !      220 MILLION        !
----------------------------------------------------------------------------------!
!                          !                            !                         !
!     0.10000000E 01       !      0.10000000E 01        !    0.99885064E 00       !
!                          !                            !                         !
!     0.10000000E 01       !      0.10000000E 01        !    0.99883879E 00       !
!                          !                            !                         !
!     0.10000000E 01       !      0.10000000E 01        !    0.99882676E 00       !
!                          !                            !                         !
!     0.10000000E 01       !      0.10000000E 01        !    0.99881455E 00       !
!                          !                            !                         !
!     0.10000000E 01       !      0.10000000E 01        !    0.99880215E 00       !
!                          !                            !                         !
!     0.10000000E 01       !      0.10000000E 01        !    0.99878955E 00       !
!                          !                            !                         !
!     0.10000000E 01       !      0.10000000E 01        !    0.99877676E 00       !
!                          !                            !                         !
!     0.10000000E 01       !      0.10000000E 01        !    0.99876377E 00       !
!                          !                            !                         !
!     0.10000000E 01       !      0.10000000E 01        !    0.99875058E 00       !
!                          !                            !                         !
!     0.10000000E 01       !      0.10000000E 01        !    0.99873717E 00       !
!                          !                            !                         !
!     0.10000000E 01       !      0.10000000E 01        !    0.99872355E 00       !
!                          !                            !                         !
!     0.10000000E 01       !      0.10000000E 01        !    0.99870972E 00       !
!                          !                            !                         !
!     0.10000000E 01       !      0.10000000E 01        !    0.99869566E 00       !
----------------------------------------------------------------------------------
```

Table 4b. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS DATE OF BIRTH (YEAR, MONTH, AND DAY) AND SURNAME

SURNAME DECILE 2

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.99992041E 00	0.99951156E 00	0.97087
0.50%	0.9900	0.99991933E 00	0.99950635E 00	0.97053
1.00%	0.9801	0.99991822E 00	0.99950106E 00	0.97030
1.50%	0.9702	0.99991710E 00	0.99949569E 00	0.97000
2.00%	0.9604	0.99991596E 00	0.99949023E 00	0.9697
2.50%	0.9506	0.99991480E 00	0.99948469E 00	0.96940
3.00%	0.9409	0.99991362E 00	0.99947906E 00	0.96910
3.50%	0.9312	0.99991243E 00	0.99947334E 00	0.96879
4.00%	0.9216	0.99991121E 00	0.99946754E 00	0.96847
4.50%	0.9120	0.99990997E 00	0.99946163E 00	0.96815
5.00%	0.9025	0.99990871E 00	0.99945564E 00	0.96782
5.50%	0.8930	0.99990742E 00	0.99944954E 00	0.96749
6.00%	0.8836	0.99990612E 00	0.99944335E 00	0.96710

CISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 2

FILE SIZE		
1 MILLION	4 MILLION	220 MILLION
0.99992041E 00	0.99951156E 00	0.97087717E 00
0.99991933E 00	0.99950635E 00	0.97059209E 00
0.99991822E 00	0.99950106E 00	0.97030286E 00
0.99991710E 00	0.99949569E 00	0.97000939E 00
0.99991596E 00	0.99949023E 00	0.96971159E 00
0.99991480E 00	0.99948469E 00	0.96940938E 00
0.99991362E 00	0.99947906E 00	0.96910268E 00
0.99991243E 00	0.99947334E 00	0.96879140E 00
0.99991121E 00	0.99946754E 00	0.96847544E 00
0.99990997E 00	0.99946163E 00	0.96815472E 00
0.99990871E 00	0.99945564E 00	0.96782914E 00
0.99990742E 00	0.99944954E 00	0.96749859E 00
0.99990612E 00	0.99944335E 00	0.96716299E 00

Table 4c. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 3

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.99952913E 00	0.99792813E 00	0.89469
0.50%	0.9900	0.99952407E 00	0.99790698E 00	0.89374
1.00%	0.9801	0.99951893E 00	0.99788550E 00	0.89278
1.50%	0.9702	0.99951372E 00	0.99786369E 00	0.89181
2.00%	0.9604	0.99950841E 00	0.99784155E 00	0.89082
2.50%	0.9506	0.99950303E 00	0.99781906E 00	0.88982
3.00%	0.9409	0.99949756E 00	0.99779623E 00	0.88881
3.50%	0.9312	0.99949200E 00	0.99777304E 00	0.88779
4.00%	0.9216	0.99948636E 00	0.99774948E 00	0.88675
4.50%	0.9120	0.99948062E 00	0.99772555E 00	0.88569
5.00%	0.9025	0.99947479E 00	0.99770125E 00	0.88463
5.50%	0.8930	0.99946887E 00	0.99767655E 00	0.88354
6.00%	0.8836	0.99946285E 00	0.99765146E 00	0.88245

ION OF RETRIEVED RESULTS FOR DATA FIELDS
TE OF BIRTH (YEAR, MONTH, AND DAY)
D SURNAME

SURNAME DECILE 3		
FILE SIZE		
1 MILLION	4 MILLION	220 MILLION
0.99952913E 00	0.99792813E 00	0.89465778E 00
0.99952407E 00	0.99790698E 00	0.89374927E 00
0.99951893E 00	0.99788550E 00	0.89278842E 00
0.99951372E 00	0.99786369E 00	0.89181500E 00
0.99950841E 00	0.99784155E 00	0.89082881E 00
0.99950303E 00	0.99781906E 00	0.88982964E 00
0.99949756E 00	0.99779623E 00	0.88881727E 00
0.99949200E 00	0.99777304E 00	0.88779147E 00
0.99948636E 00	0.99774948E 00	0.88675203E 00
0.99948062E 00	0.99772555E 00	0.88569869E 00
0.99947479E 00	0.99770125E 00	0.88463125E 00
0.99946887E 00	0.99767655E 00	0.88354946E 00
0.99946285E 00	0.99765146E 00	0.88245303E 00

Table 4d. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS DATE OF BIRTH (YEAR, MONTH, AND DAY) AND SURNAME

SURNAME DECILE 4

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MILL
0.00%	1.0000	0.99857248E 00	0.99407259E 00	0.7505216
0.50%	0.9900	0.99855771E 00	0.99401282E 00	0.7486395
1.00%	0.9801	0.99854272E 00	0.99395214E 00	0.7467386
1.50%	0.9702	0.99852749E 00	0.99389054E 00	0.7448184
2.00%	0.9604	0.99851203E 00	0.99382800E 00	0.7428788
2.50%	0.9506	0.99849633E 00	0.99376451E 00	0.7409197
3.00%	0.9409	0.99848039E 00	0.99370003E 00	0.7389407
3.50%	0.9312	0.99846419E 00	0.99363456E 00	0.7369417
4.00%	0.9216	0.99844774E 00	0.99356807E 00	0.7349223
4.50%	0.9120	0.99843102E 00	0.99350053E 00	0.7328825
5.00%	0.9025	0.99841405E 00	0.99343194E 00	0.7308220
5.50%	0.8930	0.99839679E 00	0.99336226E 00	0.7287405
6.00%	0.8836	0.99837926E 00	0.99329147E 00	0.7266376

```
ON OF RETRIEVED RESULTS FOR DATA FIELDS
TE OF BIRTH (YEAR, MONTH, AND DAY)
D SURNAME
-------------------------------------------------------------------------------
                                                                              !
         SURNAME DECILE  4                                                    !
------------------------------------------------------------------------------!
                              FILE SIZE                                       !
------------------------------------------------------------------------------!
                         !                           !     220 MILLION        !
                         !        4 MILLION          !------------------------!
      1 MILLION          !                           !                        !
-------------------------------------------------------------------------------
                         !                           !   0.75052162E 00       !
                         !     0.99407259E 00        !                        !
  0.99857248E 00         !                           !   0.74863958E 00       !
                         !     0.99401282E 00        !                        !
  0.99855771E 00         !                           !   0.74673860E 00       !
                         !     0.99395214E 00        !                        !
  0.99854272E 00         !                           !   0.74481844E 00       !
                         !     0.99389054E 00        !                        !
  0.99852749E 00         !                           !   0.74287989E 00       !
                         !     0.99382800E 00        !                        !
  0.99851203E 00         !                           !   0.74091974E 00       !
                         !     0.99376451E 00        !                        !
  0.99849633E 00         !                           !   0.73894076E 00       !
                         !     0.99370003E 00        !                        !
  0.99848039E 00         !                           !   0.73694173E 00       !
                         !     0.99363456E 00        !                        !
  0.99846419E 00         !                           !   0.73492239E 00       !
                         !     0.99356807E 00        !                        !
  0.99844774E 00         !                           !   0.73288257E 00       !
                         !     0.99350053E 00        !                        !
  0.99843102E 00         !                           !   0.73082203E 00       !
                         !     0.99343194E 00        !                        !
  0.99841405E 00         !                           !   0.72874053E 00       !
                         !     0.99336226E 00        !                        !
  0.99839679E 00         !                           !   0.72663701E 00       !
                         !     0.99329147E 00        !------------------------
  0.99837926E 00         !----------------------------
-------------------------
```

Table 4e. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS DATE OF BIRTH (YEAR, MONTH, AND DAY) AND SURNAME

SURNAME DECILE 5

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.99743178E 00	0.98950448E 00	0.62934
0.50%	0.9900	0.99740546E 00	0.98939934E 00	0.62699
1.00%	0.9801	0.99737873E 00	0.98929263E 00	0.62463
1.50%	0.9702	0.99735160E 00	0.98918431E 00	0.62226
2.00%	0.9604	0.99732405E 00	0.98907434E 00	0.61986
2.50%	0.9506	0.99729608E 00	0.98896271E 00	0.61745
3.00%	0.9409	0.99726767E 00	0.98884936E 00	0.61502
3.50%	0.9312	0.99723882E 00	0.98873427E 00	0.61256
4.00%	0.9216	0.99720951E 00	0.98861741E 00	0.61010
4.50%	0.9120	0.99717975E 00	0.98849873E 00	0.60761
5.00%	0.9025	0.99714950E 00	0.98837820E 00	0.60510
5.50%	0.8930	0.99711878E 00	0.98825577E 00	0.60258
6.00%	0.8836	0.99708756E 00	0.98813142E 00	0.6000

```
SION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR. MONTH. AND DAY)
AND SURNAME
-----------------------------------------------------------------------------
                                                                            !
        SURNAME DECILE  5                                                   !
----------------------------------------------------------------------------!
                              FILE SIZE                                     !
----------------------------------------------------------------------------!
      1 MILLION        !          4 MILLION          !      220 MILLION     !
----------------------------------------------------------------------------!
                       !                             !                      !
    0.99743178E 00     !       0.98950448E 00        !    0.62934034E 00    !
                       !                             !                      !
    0.99740546E 00     !       0.98939934E 00        !    0.62699853E 00    !
                       !                             !                      !
    0.99737873E 00     !       0.98929263E 00        !    0.62463895E 00    !
                       !                             !                      !
    0.99735160E 00     !       0.98918431E 00        !    0.62226144E 00    !
                       !                             !                      !
    0.99732405E 00     !       0.98907434E 00        !    0.61986586E 00    !
                       !                             !                      !
    0.99729608E 00     !       0.98896271E 00        !    0.61745216E 00    !
                       !                             !                      !
    0.99726767E 00     !       0.98884936E 00        !    0.61502021E 00    !
                       !                             !                      !
    0.99723882E 00     !       0.98873427E 00        !    0.61256985E 00    !
                       !                             !                      !
    0.99720951E 00     !       0.98861741E 00        !    0.61010101E 00    !
                       !                             !                      !
    0.99717975E 00     !       0.98849873E 00        !    0.60761360E 00    !
                       !                             !                      !
    0.99714950E 00     !       0.98837820E 00        !    0.60510752E 00    !
                       !                             !                      !
    0.99711878E 00     !       0.98825577E 00        !    0.60258260E 00    !
                       !                             !                      !
    0.99708756E 00     !       0.98813142E 00        !    0.60003877E 00    !
-----------------------------------------------------------------------------
```

Table 4f. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS DATE OF BIRTH (YEAR, MONTH, AND DAY) AND SURNAME

SURNAME DECILE 6

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.99645179E 00	0.98560510E 00	0.55257
0.50%	0.9900	0.99641557E 00	0.98546104E 00	0.55005
1.00%	0.9801	0.99637880E 00	0.98531597E 00	0.54755
1.50%	0.9702	0.99634146E 00	0.98516813E 00	0.54504
2.00%	0.9604	0.99630355E 00	0.98501806E 00	0.54252
2.50%	0.9506	0.99626506E 00	0.98486572E 00	0.53998
3.00%	0.9409	0.99622597E 00	0.98471106E 00	0.53742
3.50%	0.9312	0.99618627E 00	0.98455403E 00	0.53485
4.00%	0.9216	0.99614595E 00	0.98439460E 00	0.53227
4.50%	0.9120	0.99610499E 00	0.98423270E 00	0.52966
5.00%	0.9025	0.99606339E 00	0.98406829E 00	0.52705
5.50%	0.8930	0.99602112E 00	0.98390132E 00	0.52442
6.00%	0.8836	0.99597818E 00	0.98373173E 00	0.52177

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 DATE OF BIRTH (YEAR, MONTH, AND DAY)
 AND SURNAME
---------------------------------------------------------------------------
!
!        SURNAME DECILE  6                                                 !
---------------------------------------------------------------------------
!                              FILE SIZE                                   !
!--------------------------------------------------------------------------
!        1 MILLION        !        4 MILLION        !      220 MILLION     !
---------------------------------------------------------------------------
!                         !                         !                      !
!    0.99645179E 00       !    0.98560510E 00       !   0.55253374E 00     !
!    0.99641557E 00       !    0.98546164E 00       !   0.55005364E 00     !
!    0.99637880E 00       !    0.98531597E 00       !   0.54755860E 00     !
!    0.99634146E 00       !    0.98516813E 00       !   0.54504847E 00     !
!    0.99630355E 00       !    0.98501806E 00       !   0.54252328E 00     !
!    0.99626506E 00       !    0.98486572E 00       !   0.53998297E 00     !
!    0.99622597E 00       !    0.98471106E 00       !   0.53742749E 00     !
!    0.99618627E 00       !    0.98455403E 00       !   0.53485684E 00     !
!    0.99614595E 00       !    0.98439460E 00       !   0.53227093E 00     !
!    0.99610499E 00       !    0.98423270E 00       !   0.52966980E 00     !
!    0.99606339E 00       !    0.98406829E 00       !   0.52705340E 00     !
!    0.99602112E 00       !    0.98390132E 00       !   0.52442171E 00     !
!    0.99597818E 00       !    0.98373173E 00       !   0.52177464E 00     !
---------------------------------------------------------------------------
```

Table 4g. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS DATE OF BIRTH (YEAR, MONTH, AND DAY) AND SURNAME

SURNAME DECILE 7				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.99319994E 00	0.97282956E 00	0.39278
0.50%	0.9900	0.99313090E 00	0.97256245E 00	0.39033
1.00%	0.9801	0.99306091E 00	0.97229142E 00	0.38800
1.50%	0.9702	0.99298987E 00	0.97201642E 00	0.38559
2.00%	0.9604	0.99291773E 00	0.97173734E 00	0.38318
2.50%	0.9506	0.99284450E 00	0.97145412E 00	0.38077
3.00%	0.9409	0.99277013E 00	0.97116667E 00	0.37835
3.50%	0.9312	0.99269461E 00	0.97087491E 00	0.37592
4.00%	0.9216	0.99261792E 00	0.97057875E 00	0.37349
4.50%	0.9120	0.99254002E 00	0.97027812E 00	0.37104
5.00%	0.9025	0.99246090E 00	0.96997290E 00	0.36860
5.50%	0.8930	0.99238052E 00	0.96966302E 00	0.36614
6.00%	0.8836	0.99229887E 00	0.96934839E 00	0.36369

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 DATE OF BIRTH (YEAR, MONTH, AND DAY)
 AND SURNAME
-------------------------------------------------------------------------------
                                                                              !
        SURNAME DECILE  7                                                     !
-------------------------------------------------------------------------------
!                                FILE SIZE                                    !
!-----------------------------------------------------------------------------!
!         1 MILLION        !          4 MILLION          !     220 MILLION    !
-------------------------------------------------------------------------------
!                          !                             !                    !
!      0.99319984E 00      !       0.97282956E 00        !   0.39278409E 00   !
!                          !                             !                    !
!      0.99313090E 00      !       0.97256245E 00        !   0.39039549E 00   !
!                          !                             !                    !
!      0.99306091E 00      !       0.97229142E 00        !   0.38800017E 00   !
!                          !                             !                    !
!      0.99298987E 00      !       0.97201642E 00        !   0.38559817E 00   !
!                          !                             !                    !
!      0.99291773E 00      !       0.97173734E 00        !   0.38318955E 00   !
!                          !                             !                    !
!      0.99284450E 00      !       0.97145412E 00        !   0.38077432E 00   !
!                          !                             !                    !
!      0.99277013E 00      !       0.97116667E 00        !   0.37835265E 00   !
!                          !                             !                    !
!      0.99269461E 00      !       0.97087491E 00        !   0.37592456E 00   !
!                          !                             !                    !
!      0.99261792E 00      !       0.97057875E 00        !   0.37349012E 00   !
!                          !                             !                    !
!      0.99254002E 00      !       0.97027812E 00        !   0.37104939E 00   !
!                          !                             !                    !
!      0.99246090E 00      !       0.96997290E 00        !   0.36860252E 00   !
!                          !                             !                    !
!      0.99238052E 00      !       0.96966302E 00        !   0.36614950E 00   !
!                          !                             !                    !
!      0.99229887E 00      !       0.96934839E 00        !   0.36369045E 00   !
-------------------------------------------------------------------------------
```

Table 4h. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 8

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MILL
0.00%	1.0000	0.97443133E 00	0.90369576E 00	0.1451336
0.50%	0.9900	0.97417780E 00	0.90281755E 00	0.1438941
1.00%	0.9801	0.97392053E 00	0.90192770E 00	0.1426573
1.50%	0.9702	0.97365947E 00	0.90102607E 00	0.1414232
2.00%	0.9604	0.97339453E 00	0.90011242E 00	0.1401918
2.50%	0.9506	0.97312564E 00	0.89918657E 00	0.1389631
3.00%	0.9409	0.97285272E 00	0.89824830E 00	0.1377373
3.50%	0.9312	0.97257568E 00	0.89729739E 00	0.1365143
4.00%	0.9216	0.97229446E 00	0.89633364E 00	0.1352941
4.50%	0.9120	0.97200895E 00	0.89535682E 00	0.1340769
5.00%	0.9025	0.97171909E 00	0.89436669E 00	0.1328626
5.50%	0.8930	0.97142477E 00	0.89336303E 00	0.1316513
6.00%	0.8836	0.97112591E 00	0.89234561E 00	0.1304431

ION OF RETRIEVED RESULTS FOR DATA FIELDS
ATE OF BIRTH (YEAR, MONTH, AND DAY)
ND SURNAME

· SURNAME DECILE 8

FILE SIZE		
1 MILLION	4 MILLION	220 MILLION
0.97443133E 00	0.90369576E 00	0.14513361E 00
0.97417780E 00	0.90281755E 00	0.14389416E 00
0.97392053E 00	0.90192770E 00	0.14265735E 00
0.97365947E 00	0.90102607E 00	0.14142324E 00
0.97339453E 00	0.90011242E 00	0.14019184E 00
0.97312564E 00	0.89918657E 00	0.13896318E 00
0.97285272E 00	0.89824830E 00	0.13773733E 00
0.97257568E 00	0.89729739E 00	0.13651431E 00
0.97229446E 00	0.89633364E 00	0.13530415E 00
0.97200895E 00	0.89535682E 00	0.13407693E 00
0.97171909E 00	0.89436669E 00	0.13286265E 00
0.97142477E 00	0.89336303E 00	0.13165137E 00
0.97112591E 00	0.89234561E 00	0.13044311E 00

Table 4i. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS DATE OF BIRTH (YEAR, MONTH, AND DAY) AND SURNAME

SURNAME DECILE 9

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.95907313E 00	0.85236535E 00	0.94593
0.50%	0.9900	0.95867393E 00	0.85109628E 00	0.93737
1.00%	0.9801	0.95826898E 00	0.84981179E 00	0.92885
1.50%	0.9702	0.95785810E 00	0.84851162E 00	0.92039
2.00%	0.9604	0.95744142E 00	0.84719553E 00	0.91188
2.50%	0.9506	0.95701050E 00	0.84586333E 00	0.90344
3.00%	0.9409	0.95658954E 00	0.84451472E 00	0.89502
3.50%	0.9312	0.95615419E 00	0.84314946E 00	0.88664
4.00%	0.9216	0.95571241E 00	0.84176734E 00	0.87828
4.50%	0.9120	0.95526407E 00	0.84036806E 00	0.86994
5.00%	0.9025	0.95480303E 00	0.83895135E 00	0.86164
5.50%	0.8930	0.95434717E 00	0.83751697E 00	0.85337
6.00%	0.8836	0.95387036E 00	0.83606462E 00	0.84512

```
SION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME
----------------------------------------------------------------------------
                                                                           !
      SURNAME DECILE  9                                                    !
---------------------------------------------------------------------------!
                           FILE SIZE                                       !
---------------------------------------------------------------------------!
      1 MILLION        !          4 MILLION         !      220 MILLION     !
---------------------------------------------------------------------------!
                       !                            !                      !
    0.95907313E 00     !       0.85236535E 00       !    0.94593066E-01    !
                       !                            !                      !
    0.95867393E 00     !       0.85109628E 00       !    0.93737908E-01    !
                       !                            !                      !
    0.95826898E 00     !       0.84981179E 00       !    0.92885439E-01    !
                       !                            !                      !
    0.95785810E 00     !       0.84851162E 00       !    0.92035667E-01    !
                       !                            !                      !
    0.95744142E 00     !       0.84719553E 00       !    0.91188621E-01    !
                       !                            !                      !
    0.95701858E 00     !       0.84586333E 00       !    0.90344316E-01    !
                       !                            !                      !
    0.95658954E 00     !       0.84451472E 00       !    0.89502771E-01    !
                       !                            !                      !
    0.95615419E 00     !       0.84314946E 00       !    0.88664000E-01    !
                       !                            !                      !
    0.95571241E 00     !       0.84176734E 00       !    0.87828036E-01    !
                       !                            !                      !
    0.95526407E 00     !       0.84036806E 00       !    0.86994889E-01    !
                       !                            !                      !
    0.95480903E 00     !       0.83895135E 00       !    0.86164591E-01    !
                       !                            !                      !
    0.95434717E 00     !       0.83751697E 00       !    0.85337146E-01    !
                       !                            !                      !
    0.95387836E 00     !       0.83606462E 00       !    0.84512574E-01    !
----------------------------------------------------------------------------
```

Table 4j. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS DATE OF BIRTH (YEAR, MONTH, AND DAY) AND SURNAME

SURNAME DECILE 10

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.84419429E 00	0.57189911E 90	0.2360
0.50%	0.9900	0.84285081E 00	0.56943730E 00	0.2337
1.00%	0.9801	0.84150720E 00	0.56695962E 00	0.2315
1.50%	0.9702	0.84013948E 00	0.56446605E 00	0.2292
2.00%	0.9604	0.83875517E 00	0.56195653E 00	0.2269
2.50%	0.9506	0.83735413E 00	0.55943106E 00	0.2246
3.00%	0.9409	0.83593608E 00	0.55688947E 00	0.2224
3.50%	0.9312	0.83450074E 00	0.55433178E 00	0.2202
4.00%	0.9216	0.83304790E 00	0.55175794E 00	0.2179
4.50%	0.9120	0.83157729E 00	0.54916790E 00	0.2157
5.00%	0.9025	0.83008863E 00	0.54656150E 00	0.2135
5.50%	0.8930	0.82858166E 00	0.54393898E 00	0.2113
6.00%	0.8836	0.82705609E 00	0.54130002E 00	0.2091

ISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 10

FILE SIZE		
1 MILLION	4 MILLION	220 MILLION
0.84419429E 00	0.57189911E 00	0.23609111E-01
0.84285081E 00	0.56943730E 00	0.23379105E-01
0.84150728E 00	0.56695962E 00	0.23150145E-01
0.84013948E 00	0.56446605E 00	0.22922236E-01
0.83875517E 00	0.56195653E 00	0.22695373E-01
0.83735413E 00	0.55943106E 00	0.22469559E-01
0.83593608E 00	0.55688947E 00	0.22244797E-01
0.83450074E 00	0.55433178E 00	0.22021089E-01
0.83304790E 00	0.55175794E 00	0.21799434E-01
0.83157729E 00	0.54916790E 00	0.21576836E-01
0.83008863E 00	0.54656158E 00	0.21356298E-01
0.82858166E 00	0.54393898E 00	0.21136817E-01
0.82705609E 00	0.54130002E 00	0.20918397E-01

Table 5a. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
STREET NAME
STREET NUMBER
AND SURNAME

	SURNAME DECILE			
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.1000
0.50%	0.9801	0.10000000E 01	0.10000000E 01	0.1000
1.00%	0.9606	0.10000000E 01	0.10000000E 01	0.1000
1.50%	0.9413	0.10000000E 01	0.10000000E 01	0.1000
2.00%	0.9224	0.10000000E 01	0.10000000E 01	0.1000
2.50%	0.9037	0.10000000E 01	0.10000000E 01	0.1000
3.00%	0.8853	0.10000000E 01	0.10000000E 01	0.1000
3.50%	0.8672	0.10000000E 01	0.10000000E 01	0.1000
4.00%	0.8493	0.10000000E 01	0.10000000E 01	0.1000
4.50%	0.8318	0.10000000E 01	0.10000000E 01	0.1000
5.00%	0.8145	0.10000000E 01	0.10000000E 01	0.1000
5.50%	0.7975	0.10000000E 01	0.10000000E 01	0.1000
6.00%	0.7807	0.10000000E 01	0.10000000E 01	0.1000

```
SION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
STREET NAME
STREET NUMBER
AND SURNAME
-----------------------------------------------------------------------------------------
                                                                                         !
      SURNAME DECILE  1                                                                  !
-----------------------------------------------------------------------------------------!
                                   FILE SIZE                                             !
-----------------------------------------------------------------------------------------!
      1 MILLION          !           4 MILLION            !        220 MILLION           !
-----------------------------------------------------------------------------------------!
                         !                                !                              !
   0.10000000E 01        !        0.10000000E 01          !      0.10000000E 01          !
                         !                                !                              !
   0.10000000E 01        !        0.10000000E 01          !      0.10000000E 01          !
                         !                                !                              !
   0.10000000E 01        !        0.10000000E 01          !      0.10000000E 01          !
                         !                                !                              !
   0.10000000E 01        !        0.10000000E 01          !      0.10000000E 01          !
                         !                                !                              !
   0.10000000E 01        !        0.10000000E 01          !      0.10000000E 01          !
                         !                                !                              !
   0.10000000E 01        !        0.10000000E 01          !      0.10000000E 01          !
                         !                                !                              !
   0.10000000E 01        !        0.10000000E 01          !      0.10000000E 01          !
                         !                                !                              !
   0.10000000E 01        !        0.10000000E 01          !      0.10000000E 01          !
                         !                                !                              !
   0.10000000E 01        !        0.10000000E 01          !      0.10000000E 01          !
                         !                                !                              !
   0.10000000E 01        !        0.10000000E 01          !      0.10000000E 01          !
                         !                                !                              !
   0.10000000E 01        !        0.10000000E 01          !      0.10000000E 01          !
                         !                                !                              !
   0.10000000E 01        !        0.10000000E 01          !      0.10000000E 01          !
                         !                                !                              !
   0.10000000E 01        !        0.10000000E 01          !      0.10000000E 01          !
-----------------------------------------------------------------------------------------
```

Table 5b. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
STREET NAME
STREET NUMBER
AND SURNAME

SURNAME DECILE 2

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.1000
0.50%	0.9801	0.10000000E 01	0.10000000E 01	0.1000
1.00%	0.9606	0.10000000E 01	0.10000000E 01	0.1000
1.50%	0.9413	0.10000000E 01	0.10000000E 01	0.1000
2.00%	0.9224	0.10000000E 01	0.10000000E 01	0.1000
2.50%	0.9037	0.10000000E 01	0.10000000E 01	0.1000
3.00%	0.8853	0.10000000E 01	0.10000000E 01	0.1000
3.50%	0.8672	0.10000000E 01	0.10000000E 01	0.1000
4.00%	0.8493	0.10000000E 01	0.10000000E 01	0.1000
4.50%	0.8318	0.10000000E 01	0.10000000E 01	0.1000
5.00%	0.8145	0.10000000E 01	0.10000000E 01	0.1000
5.50%	0.7975	0.10000000E 01	0.10000000E 01	0.1000
6.00%	0.7807	0.10000000E 01	0.10000000E 01	0.1000

```
CISION OF RETRIEVED RESULTS FOR DATA FIELDS
 DATE OF BIRTH (YEAR, MONTH, AND DAY)
 STREET NAME
 STREET NUMBER
 AND SURNAME
------------------------------------------------------------------------------
          SURNAME DECILE  2
------------------------------------------------------------------------------
!                                                                            !
!                            FILE SIZE                                       !
!----------------------------------------------------------------------------!
!         1 MILLION          !        4 MILLION          !     220 MILLION   !
------------------------------------------------------------------------------
!                            !                           !                   !
!     0.10000000E 01         !     0.10000000E 01        !  0.10000000E 01   !
!     0.10000000E 01         !     0.10000000E 01        !  0.10000000E 01   !
!     0.10000000E 01         !     0.10000000E 01        !  0.10000000E 01   !
!     0.10000000E 01         !     0.10000000E 01        !  0.10000000E 01   !
!     0.10000000E 01         !     0.10000000E 01        !  0.10000000E 01   !
!     0.10000000E 01         !     0.10000000E 01        !  0.10000000E 01   !
!     0.10000000E 01         !     0.10000000E 01        !  0.10000000E 01   !
!     0.10000000E 01         !     0.10000000E 01        !  0.10000000E 01   !
!     0.10000000E 01         !     0.10000000E 01        !  0.10000000E 01   !
!     0.10000000E 01         !     0.10000000E 01        !  0.10000000E 01   !
!     0.10000000E 01         !     0.10000000E 01        !  0.10000000E 01   !
!     0.10000000E 01         !     0.10000000E 01        !  0.10000000E 01   !
!     0.10000000E 01         !     0.10000000E 01        !  0.10000000E 01   !
------------------------------------------------------------------------------
```

Table 5c. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
STREET NAME
STREET NUMBER
AND SURNAME

SURNAME DECILE 3				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.10000
0.50%	0.9801	0.10000000E 01	0.10000000E 01	0.10000
1.00%	0.9606	0.10000000E 01	0.10000000E 01	0.10000
1.50%	0.9413	0.10000000E 01	0.10000000E 01	0.10000
2.00%	0.9224	0.10000000E 01	0.10000000E 01	0.10000
2.50%	0.9037	0.10000000E 01	0.10000000E 01	0.10000
3.00%	0.8853	0.10000000E 01	0.10000000E 01	0.10000
3.50%	0.8672	0.10000000E 01	0.10000000E 01	0.10000
4.00%	0.8493	0.10000000E 01	0.10000000E 01	0.10000
4.50%	0.8318	0.10000000E 01	0.10000000E 01	0.10000
5.00%	0.8145	0.10000000E 01	0.10000000E 01	0.10000
5.50%	0.7975	0.10000000E 01	0.10000000E 01	0.10000
6.00%	0.7807	0.10000000E 01	0.10000000E 01	0.10000

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 DATE OF BIRTH (YEAR, MONTH, AND DAY)
 STREET NAME
 STREET NUMBER
 AND SURNAME
---------------------------------------------------------------------------
         SURNAME DECILE  3
---------------------------------------------------------------------------
                           FILE SIZE
---------------------------------------------------------------------------
       1 MILLION        !        4 MILLION         !       220 MILLION      !
---------------------------------------------------------------------------
                        !                          !                        !
     0.10000000E 01     !      0.10000000E 01      !     0.10000000E 01     !
     0.10000000E 01     !      0.10000000E 01      !     0.10000000E 01     !
     0.10000000E 01     !      0.10000000E 01      !     0.10000000E 01     !
     0.10000000E 01     !      0.10000000E 01      !     0.10000000E 01     !
     0.10000000E 01     !      0.10000000E 01      !     0.10000000E 01     !
     0.10000000E 01     !      0.10000000E 01      !     0.10000000E 01     !
     0.10000000E 01     !      0.10000000E 01      !     0.10000000E 01     !
     0.10000000E 01     !      0.10000000E 01      !     0.10000000E 01     !
     0.10000000E 01     !      0.10000000E 01      !     0.10000000E 01     !
     0.10000000E 01     !      0.10000000E 01      !     0.10000000E 01     !
     0.10000000E 01     !      0.10000000E 01      !     0.10000000E 01     !
     0.10000000E 01     !      0.10000000E 01      !     0.10000000E 01     !
     0.10000000E 01     !      0.10000000E 01      !     0.10000000E 01     !
---------------------------------------------------------------------------
```

Table 5d. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
STREET NAME
STREET NUMBER
AND SURNAME

SURNAME DECILE 4

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.99999
0.50%	0.9801	0.10000000E 01	0.10000000E 01	0.99999
1.00%	0.9606	0.10000000E 01	0.10000000E 01	0.99999
1.50%	0.9413	0.10000000E 01	0.10000000E 01	0.99999
2.00%	0.9224	0.10000000E 01	0.10000000E 01	0.99999
2.50%	0.9037	0.10000000E 01	0.10000000E 01	0.99999
3.00%	0.8853	0.10000000E 01	0.10000000E 01	0.99999
3.50%	0.8672	0.10000000E 01	0.10000000E 01	0.99999
4.00%	0.8493	0.10000000E 01	0.10000000E 01	0.99999
4.50%	0.8318	0.10000000E 01	0.10000000E 01	0.99999
5.00%	0.8145	0.10000000E 01	0.10000000E 01	0.99999
5.50%	0.7975	0.10000000E 01	0.10000000E 01	0.99999
6.00%	0.7807	0.10000000E 01	0.10000000E 01	0.99999

```
CISION OF RETRIEVED RESULTS FOR DATA FIELDS
  DATE OF BIRTH (YEAR, MONTH, AND DAY)
  STREET NAME
  STREET NUMBER
  AND SURNAME
-----------------------------------------------------------------------
                                                                      !
         SURNAME DECILE  4                                            !
-----------------------------------------------------------------------
!                             FILE SIZE                               !
!---------------------------------------------------------------------!
!      1 MILLION       !        4 MILLION        !     220 MILLION    !
-----------------------------------------------------------------------
!                      !                         !                    !
!   0.10000000E 01     !     0.10000000E 01      !   0.99999999E 00   !
!                      !                         !                    !
!   0.10000000E 01     !     0.10000000E 01      !   0.99999999E 00   !
!                      !                         !                    !
!   0.10000000E 01     !     0.10000000E 01      !   0.99999999E 00   !
!                      !                         !                    !
!   0.10000000E 01     !     0.10000000E 01      !   0.99999999E 00   !
!                      !                         !                    !
!   0.10000000E 01     !     0.10000000E 01      !   0.99999999E 00   !
!                      !                         !                    !
!   0.10000000E 01     !     0.10000000E 01      !   0.99999999E 00   !
!                      !                         !                    !
!   0.10000000E 01     !     0.10000000E 01      !   0.99999999E 00   !
!                      !                         !                    !
!   0.10000000E 01     !     0.10000000E 01      !   0.99999999E 00   !
!                      !                         !                    !
!   0.10000000E 01     !     0.10000000E 01      !   0.99999999E 00   !
!                      !                         !                    !
!   0.10000000E 01     !     0.10000000E 01      !   0.99999999E 00   !
!                      !                         !                    !
!   0.10000000E 01     !     0.10000000E 01      !   0.99999999E 00   !
!                      !                         !                    !
!   0.10000000E 01     !     0.10000000E 01      !   0.99999999E 00   !
!                      !                         !                    !
!   0.10000000E 01     !     0.10000000E 01      !   0.99999999E 00   !
-----------------------------------------------------------------------
```

Table 5e. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
STREET NAME
STREET NUMBER
AND SURNAME

SURNAME DECILE 5

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.99999
0.50%	0.9801	0.10000000E 01	0.10000000E 01	0.99999
1.00%	0.9606	0.10000000E 01	0.10000000E 01	0.99999
1.50%	0.9413	0.10000000E 01	0.10000000E 01	0.99999
2.00%	0.9224	0.10000000E 01	0.10000000E 01	0.99999
2.50%	0.9037	0.10000000E 01	0.10000000E 01	0.99999
3.00%	0.8853	0.10000000E 01	0.10000000E 01	0.99999
3.50%	0.8672	0.10000000E 01	0.10000000E 01	0.99999
4.00%	0.8493	0.10000000E 01	0.10000000E 01	0.99999
4.50%	0.8310	0.10000000E 01	0.10000000E 01	0.99999
5.00%	0.8145	0.10000000E 01	0.10000000E 01	0.99999
5.50%	0.7975	0.10000000E 01	0.10000000E 01	0.99999
6.00%	0.7807	0.10000000E 01	0.10000000E 01	0.99999

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 DATE OF BIRTH (YEAR, MONTH, AND DAY)
 STREET NAME
 STREET NUMBER
 AND SURNAME
------------------------------------------------------------------------------
                                                                             !
         SURNAME DECILE  5                                                   !
------------------------------------------------------------------------------
!                               FILE SIZE                                    !
!----------------------------------------------------------------------------!
!         1 MILLION          !          4 MILLION          !   220 MILLION   !
------------------------------------------------------------------------------
!                            !                             !                 !
!      0.10000000E 01        !       0.10000000E 01        !  0.99999999E 00 !
!                            !                             !                 !
!      0.10000000E 01        !       0.10000000E 01        !  0.99999999E 00 !
!                            !                             !                 !
!      0.10000000E 01        !       0.10000000E 01        !  0.99999999E 00 !
!                            !                             !                 !
!      0.10000000E 01        !       0.10000000E 01        !  0.99999999E 00 !
!                            !                             !                 !
!      0.10000000E 01        !       0.10000000E 01        !  0.99999999E 00 !
!                            !                             !                 !
!      0.10000000E 01        !       0.10000000E 01        !  0.99999999E 00 !
!                            !                             !                 !
!      0.10000000E 01        !       0.10000000E 01        !  0.99999999E 00 !
!                            !                             !                 !
!      0.10000000E 01        !       0.10000000E 01        !  0.99999999E 00 !
!                            !                             !                 !
!      0.10000000E 01        !       0.10000000E 01        !  0.99999999E 00 !
!                            !                             !                 !
!      0.10000000E 01        !       0.10000000E 01        !  0.99999999E 00 !
!                            !                             !                 !
!      0.10000000E 01        !       0.10000000E 01        !  0.99999999E 00 !
!                            !                             !                 !
!      0.10000000E 01        !       0.10000000E 01        !  0.99999999E 00 !
!                            !                             !                 !
!      0.10000000E 01        !       0.10000000E 01        !  0.99999998E 00 !
------------------------------------------------------------------------------
```

Table 5f. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
STREET NAME
STREET NUMBER
AND SURNAME

SURNAME DECILE 6

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.99999
0.50%	0.9801	0.10000000E 01	0.10000000E 01	0.99999
1.00%	0.9606	0.10000000E 01	0.10000000E 01	0.99999
1.50%	0.9413	0.10000000E 01	0.10000000E 01	0.99999
2.00%	0.9224	0.10000000E 01	0.10000000E 01	0.99999
2.50%	0.9037	0.10000000E 01	0.10000000E 01	0.99999
3.00%	0.8853	0.10000000E 01	0.10000000E 01	0.99999
3.50%	0.8672	0.10000000E 01	0.10000000E 01	0.99999
4.00%	0.8493	0.10000000E 01	0.10000000E 01	0.99999
4.50%	0.8318	0.10000000E 01	0.10000000E 01	0.99999
5.00%	0.8145	0.10000000E 01	0.10000000E 01	0.99999
5.50%	0.7975	0.10000000E 01	0.10000000E 01	0.99999
6.00%	0.7807	0.10000000E 01	0.10000000E 01	0.99999

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 DATE OF BIRTH (YEAR, MONTH, AND DAY)
 STREET NAME
 STREET NUMBER
 AND SURNAME
---------------------------------------------------------------------------
                                                                          !
        SURNAME DECILE  6                                                 !
--------------------------------------------------------------------------!
!                              FILE SIZE                                  !
!-------------------------------------------------------------------------!
!       1 MILLION        !        4 MILLION         !     220 MILLION     !
--------------------------------------------------------------------------!
!                        !                          !                     !
!     0.10000000E 01     !      0.10000000E 01      !   0.99999998E 00    !
!                        !                          !                     !
!     0.10000000E 01     !      0.10000000E 01      !   0.99999998E 00    !
!                        !                          !                     !
!     0.10000000E 01     !      0.10000000E 01      !   0.99999998E 00    !
!                        !                          !                     !
!     0.10000000E 01     !      0.10000000E 01      !   0.99999998E 00    !
!                        !                          !                     !
!     0.10000000E 01     !      0.10000000E 01      !   0.99999998E 00    !
!                        !                          !                     !
!     0.10000000E 01     !      0.10000000E 01      !   0.99999998E 00    !
!                        !                          !                     !
!     0.10000000E 01     !      0.10000000E 01      !   0.99999998E 00    !
!                        !                          !                     !
!     0.10000000E 01     !      0.10000000E 01      !   0.99999998E 00    !
!                        !                          !                     !
!     0.10000000E 01     !      0.10000000E 01      !   0.99999998E 00    !
!                        !                          !                     !
!     0.10000000E 01     !      0.10000000E 01      !   0.99999998E 00    !
!                        !                          !                     !
!     0.10000000E 01     !      0.10000000E 01      !   0.99999998E 00    !
!                        !                          !                     !
!     0.10000000E 01     !      0.10000000E 01      !   0.99999998E 00    !
!                        !                          !                     !
!     0.10000000E 01     !      0.10000000E 01      !   0.99999998E 00    !
---------------------------------------------------------------------------
```

Table 5g. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
STREET NAME
STREET NUMBER
AND SURNAME

SURNAME DECILE 7				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.9999
0.50%	0.9801	0.10000000E 01	0.10000000E 01	0.9999
1.00%	0.9606	0.10000000E 01	0.10000000E 01	0.9999
1.50%	0.9413	0.10000000E 01	0.10000000E 01	0.9999
2.00%	0.9224	0.10000000E 01	0.10000000E 01	0.9999
2.50%	0.9037	0.10000000E 01	0.10000000E 01	0.9999
3.00%	0.8853	0.10000000E 01	0.10000000E 01	0.9999
3.50%	0.8672	0.10000000E 01	0.10000000E 01	0.9999
4.00%	0.8493	0.10000000E 01	0.10000000E 01	0.9999
4.50%	0.8318	0.10000000E 01	0.10000000E 01	0.9999
5.00%	0.8145	0.10000000E 01	0.10000000E 01	0.9999
5.50%	0.7975	0.10000000E 01	0.10000000E 01	0.9999
6.00%	0.7807	0.10000000E 01	0.10000000E 01	0.9999

```
SION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
STREET NAME
STREET NUMBER
AND SURNAME
-------------------------------------------------------------------------------
                                                                              !
      SURNAME DECILE  7                                                       !
-------------------------------------------------------------------------------!
                               FILE SIZE                                      !
-------------------------------------------------------------------------------!
        1 MILLION        !          4 MILLION          !      229 MILLION     !
-------------------------------------------------------------------------------!
                         !                             !                      !
     0.10000000E 01      !       0.10000000E 01        !   0.99999997E 00     !
     0.10000000E 01      !       0.10000000E 01        !   0.99999997E 00     !
     0.10000000E 01      !       0.10000000E 01        !   0.99999997E 00     !
     0.10000000E 01      !       0.10000000E 01        !   0.99999997E 00     !
     0.10000000E 01      !       0.10000000E 01        !   0.99999997E 00     !
!    0.10000000E 01      !       0.10000000E 01        !   0.99999997E 00     !
!    0.10000000E 01      !       0.10000000E 01        !   0.99999997E 00     !
!    0.10000900E 01      !       0.10000000E 01        !   0.99999996E 00     !
!    0.10000000E 01      !       0.10000000E 01        !   0.99999996E 00     !
!    0.10000000E 01      !       0.19000000E 01        !   0.99999996E 00     !
!    0.10000000E 01      !       0.10000000E 01        !   0.99999996E 00     !
!    0.10000000E 01      !       0.10000000E 01        !   0.99999996E 00     !
!    0.10000000E 01      !       0.10000000E 01        !   0.99999996E 00     !
-------------------------------------------------------------------------------
```

Table 5h. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS

DATE OF BIRTH (YEAR, MONTH, AND DAY)
STREET NAME
STREET NUMBER
AND SURNAME

SURNAME DECILE 8

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.99999
0.50%	0.9801	0.10000000E 01	0.10000000E 01	0.99999
1.00%	0.9606	0.10000000E 01	0.10000000E 01	0.99999
1.50%	0.9413	0.10000000E 01	0.10000000E 01	0.99999
2.00%	0.9224	0.10000000E 01	0.10000000E 01	0.99999
2.50%	0.9037	0.10000000E 01	0.10000000E 01	0.99999
3.00%	0.8853	0.10000000E 01	0.10000000E 01	0.99999
3.50%	0.8672	0.10000000E 01	0.10000000E 01	0.99999
4.00%	0.8493	0.10000000E 01	0.10000000E 01	0.99999
4.50%	0.8318	0.10000000E 01	0.10000000E 01	0.99999
5.00%	0.8145	0.10000000E 01	0.10000000[illegible] 01	0.99999
5.50%	0.7975	0.10000000E 01	0.1000000[illegible]	0.99999
6.00%	0.7807	0.10000000E 01	0.10000000E 01	0.99999

7

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 DATE OF BIRTH (YEAR, MONTH, AND DAY)
 STREET NAME
 STREET NUMBER
 AND SURNAME
--------------------------------------------------------------------------------
                                                                               !
         SURNAME DECILE  8                                                     !
-------------------------------------------------------------------------------!
!                            FILE SIZE                                         !
!------------------------------------------------------------------------------!
!      1 MILLION           !          4 MILLION          !    220 MILLION      !
-------------------------------------------------------------------------------!
!                          !                             !                     !
!    0.10000000E 01        !       0.10000000E 01        !   0.99999988E 00    !
!                          !                             !                     !
!    0.10000000E 01        !       0.10000000E 01        !   0.99999988E 00    !
!                          !                             !                     !
!    0.10000000E 01        !       0.10000000E 01        !   0.99999988E 00    !
!                          !                             !                     !
!    0.10000000E 01        !       0.10000000E 01        !   0.99999987E 00    !
!                          !                             !                     !
!    0.10000000E 01        !       0.10000000E 01        !   0.99999987E 00    !
!                          !                             !                     !
!    0.10000000E 01        !       0.10000000E 01        !   0.99999987E 00    !
!                          !                             !                     !
!    0.10000000E 01        !       0.10000000E 01        !   0.99999987E 00    !
!                          !                             !                     !
!    0.10000000E 01        !       0.10000000E 01        !   0.99999986E 00    !
!                          !                             !                     !
!    0.10000000E 01        !       0.10000000E 01        !   0.99999986E 00    !
!                          !                             !                     !
!    0.10000000E 01        !       0.10000000E 01        !   0.99999986E 00    !
!                          !                             !                     !
!    0.10000000E 01        !       0.10000000E 01        !   0.99999986E 00    !
!                          !                             !                     !
!    0.10000000E 01        !       0.10000000E 01        !   0.99999985E 00    !
!                          !                             !                     !
!    0.10000000E 01        !       0.10000000E 01        !   0.99999985E 00    !
--------------------------------------------------------------------------------
```

Table 5i. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
STREET NAME
STREET NUMBER
AND SURNAME

SURNAME DECILE 9

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.9999
0.50%	0.9801	0.10000000E 01	0.10000000E 01	0.9999
1.00%	0.9606	0.10000000E 01	0.10000000E 01	0.9999
1.50%	0.9413	0.10000000E 01	0.10000000E 01	0.9999
2.00%	0.9224	0.10000000E 01	0.10000000E 01	0.9999
2.50%	0.9037	0.10000000E 01	0.10000000E 01	0.9999
3.00%	0.8853	0.10000000E 01	0.10000000E 01	0.9999
3.50%	0.8672	0.10000000E 01	0.10000000E 01	0.9999
4.00%	0.8493	0.10000000E 01	0.10000000E 01	0.9999
4.50%	0.8318	0.10000000E 01	0.10000000E 01	0.9999
5.00%	0.8145	0.10000000E 01	0.10000000E 01	0.9999
5.50%	0.7975	0.10000000E 01	0.10000000E 01	0.9999
6.00%	0.7807	0.10000000E 01	0.10000000E 01	0.9999

```
ECISION OF RETRIEVED RESULTS FOR DATA FIELDS
   DATE OF BIRTH (YEAR, MONTH, AND DAY)
   STREET NAME
   STREET NUMBER
   AND SURNAME
--------------------------------------------------------------------------------
        SURNAME DECILE  9                                                      !
--------------------------------------------------------------------------------
 !                            FILE SIZE                                        !
 !-----------------------------------------------------------------------------!
 !       1 MILLION        !        4 MILLION          !      220 MILLION       !
--------------------------------------------------------------------------------
 !                        !                           !                        !
 !     0.10000000E 01     !     0.10000000E 01        !     0.99999981E 00     !
 !                        !                           !                        !
 !     0.10000000E 01     !     0.10000000E 01        !     0.99999980E 00     !
 !                        !                           !                        !
 !     0.10000000E 01     !     0.10000000E 01        !     0.99999980E 00     !
 !                        !                           !                        !
 !     0.10000000E 01     !     0.10000000E 01        !     0.99999980E 00     !
 !                        !                           !                        !
 !     0.10000000E 01     !     0.10000000E 01        !     0.99999979E 00     !
 !                        !                           !                        !
 !     0.10000000E 01     !     0.10000000E 01        !     0.99999979E 00     !
 !                        !                           !                        !
 !     0.10000000E 01     !     0.10000000E 01        !     0.99999978E 00     !
 !                        !                           !                        !
 !     0.10000000E 01     !     0.10000000E 01        !     0.99999979E 00     !
 !                        !                           !                        !
 !     0.10000000E 01     !     0.10000000E 01        !     0.99999977E 00     !
 !                        !                           !                        !
 !     0.10000000E 01     !     0.10000000E 01        !     0.99999977E 00     !
 !                        !                           !                        !
 !     0.10000000E 01     !     0.10000000E 01        !     0.99999976E 00     !
 !                        !                           !                        !
 !     0.10000000E 01     !     0.10000000E 01        !     0.99999976E 00     !
 !                        !                           !                        !
 !     0.10000000E 01     !     0.10000000E 01        !     0.99999975E 00     !
--------------------------------------------------------------------------------
```

Table 5j. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
DATE OF BIRTH (YEAR, MONTH, AND DAY)
STREET NAME
STREET NUMBER
AND SURNAME

SURNAME DECILE 10				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.10000000E 01	0.99999999E 00	0.99999
0.50%	0.9801	0.10000000E 01	0.99999998E 00	0.99999
1.00%	0.9606	0.10000000E 01	0.99999998E 00	0.99999
1.50%	0.9413	0.10000000E 01	0.99999998E 00	0.99999
2.30%	0.9224	0.10000000E 01	0.99999998E 00	0.99999
2.50%	0.9037	0.10000000E 01	0.99999998E 00	0.99999
3.00%	0.8853	0.10000000E 01	0.99999998E 00	0.99999
3.50%	0.8672	0.10000000E 01	0.99999998E 00	0.99999
4.00%	0.8493	0.10000000E 01	0.99999998E 00	0.99999
4.50%	0.8318	0.10000000E 01	0.99999998E 00	0.99999
5.00%	0.8145	0.10000000E 01	0.99999998E 00	0.9999
5.50%	0.7975	0.10000000E 01	0.99999998E 00	0.9999
6.00%	0.7807	0.10000000E 01	0.99999998E 00	0.9999

```
IEVED RESULTS FOR DATA FIELDS
H (YEAR, MONTH, AND DAY)

R

------------------------------------------------------------------------
NAME DECILE 10
-----------------------------------------------------------------------!
                 FILE SIZE                                             !
-----------------------------------------------------------------------!
LION        !          4 MILLION          !       220 MILLION          !
------------------------------------------------------------------------
            !                             !                            !
0000E 01    !       0.99999999E 00        !     0.99999917E 00         !
0000E 01    !       0.99999998E 00        !     0.99999916E 00         !
0000E 01    !       0.99999998E 00        !     0.99999914E 00         !
0000E 01    !       0.99999998E 00        !     0.99999912E 00         !
0000E 01    !       0.99999998E 00        !     0.99999910E 00         !
0000E 01    !       0.99999998E 00        !     0.99999908E 00         !
0000E 01    !       0.99999998E 00        !     0.99999907E 00         !
0000E 01    !       0.99999998E 00        !     0.99999905E 00         !
0000E 01    !       0.99999998E 00        !     0.99999903E 00         !
0000E 01    !       0.99999998E 00        !     0.99999901E 00         !
0000E 01    !       0.99999998E 00        !     0.99999898E 00         !
0000E 01    !       0.99999998E 00        !     0.99999896E 00         !
0000E 01    !       0.99999998E 00        !     0.99999894E 00         !
------------------------------------------------------------------------
```

Table 6a. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STATE
AND SURNAME

SURNAME DECILE 1

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.70707
0.50%	0.9900	0.10000000E 01	0.10000000E 01	0.79494
1.00%	0.9801	0.10000000E 01	0.10000000E 01	0.70279
1.50%	0.9702	0.10000000E 01	0.10000000E 01	0.70063
2.00%	0.9604	0.10000000E 01	0.10000000E 01	0.69843
2.50%	0.9506	0.10000000E 01	0.10000000E 01	0.69622
3.00%	0.9409	0.10000000E 01	0.10000000E 01	0.69399
3.50%	0.9312	0.10000000E 01	0.10000000E 01	0.6917
4.00%	0.9216	0.10000000E 01	0.10000000E 01	0.68946
4.50%	0.9120	0.10000000E 01	0.10000000E 01	0.68717
5.00%	0.9025	0.10000000E 01	0.10000000E 01	0.6840
5.50%	0.8930	0.10000000E 01	0.10000000E 01	0.68253
6.00%	0.8836	0.10000000E 01	0.10000000E 01	0.69017

```
CISION OF RETRIEVED RESULTS FOR DATA FIELDS
   STATE
   AND SURNAME
-------------------------------------------------------------------------
        SURNAME DECILE  1
-------------------------------------------------------------------------
 !                         FILE SIZE                                    !
 !----------------------------------------------------------------------!
 !      1 MILLION      !       4 MILLION         !     220 MILLION       !
-------------------------------------------------------------------------
 !                     !                         !                       !
 !   0.10000000E 01    !    0.10000000E 01       !   0.70707353E 00      !
 !   0.10000000E 01    !    0.10000000E 01       !   0.70494264E 00      !
 !   0.10000000E 01    !    0.10000000E 01       !   0.70279222E 00      !
 !   0.10000000E 01    !    0.10000000E 01       !   0.70062209E 00      !
 !   0.10000000E 01    !    0.10000000E 01       !   0.69843209E 00      !
 !   0.10000000E 01    !    0.10000000E 01       !   0.69622198E 00      !
 !   0.10000000E 01    !    0.10000000E 01       !   0.69399165E 00      !
 !   0.10000000E 01    !    0.10000000E 01       !   0.69171083E 00      !
 !   0.10000000E 01    !    0.10000000E 01       !   0.68946945E 00      !
 !   0.10000000E 01    !    0.10000000E 01       !   0.68717718E 00      !
 !   0.10000000E 01    !    0.10000000E 01       !   0.68488396E 00      !
 !   0.10000000E 01    !    0.10000000E 01       !   0.68252951E 00      !
 !   0.10000000E 01    !    0.10000000E 01       !   0.69017368E 00      !
-------------------------------------------------------------------------
```

Table 6b. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STATE
AND SURNAME

SURNAME DECILE 2

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.97164968E 00	0.84982140E 00	0.84748
0.50%	0.9900	0.97127591E 00	0.84846473E 00	0.83973
1.00%	0.9801	0.97089726E 00	0.84709152E 00	0.83201
1.50%	0.9702	0.97051269E 00	0.84570150E 00	0.82431
2.00%	0.9604	0.97012201E 00	0.84429444E 00	0.81663
2.50%	0.9506	0.96972511E 00	0.84287008E 00	0.80899
3.00%	0.9409	0.96932190E 00	0.84142815E 00	0.80137
3.50%	0.9312	0.96891222E 00	0.83996841E 00	0.79378
4.00%	0.9216	0.96849595E 00	0.83849059E 00	0.78621
4.50%	0.9120	0.96807295E 00	0.83699439E 00	0.77867
5.00%	0.9025	0.96764310E 00	0.83547959E 00	0.77116
5.50%	0.8930	0.96720625E 00	0.83394584E 00	0.76367
6.00%	0.8836	0.96676225E 00	0.83239291E 00	0.75622

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 STATE
 AND SURNAME
--------------------------------------------------------------------------------
                                                                               !
      SURNAME DECILE  2
-------------------------------------------------------------------------------!
!                            FILE SIZE                                         !
!------------------------------------------------------------------------------!
!       1 MILLION         !          4 MILLION          !      220 MILLION     !
-------------------------------------------------------------------------------!
!                         !                             !                      !
!      0.97164958E 00     !       0.84982140E 00        !   0.84748634E-01     !
!                         !                             !                      !
!      0.97127591E 00     !       0.84846473E 00        !   0.83973551E-01     !
!                         !                             !                      !
!      0.97089728E 00     !       0.84709152E 00        !   0.83201047E-01     !
!                         !                             !                      !
!      0.97051269E 00     !       0.84570150E 00        !   0.82431162E-01     !
!                         !                             !                      !
!      0.97012201E 00     !       0.84429444E 00        !   0.81663895E-01     !
!                         !                             !                      !
!      0.96972511E 00     !       0.84287008E 00        !   0.80899271E-01     !
!                         !                             !                      !
!      0.96932190E 00     !       0.84142815E 00        !   0.80137304E-01     !
!                         !                             !                      !
!      0.96891222E 00     !       0.83996841E 00        !   0.79378008E-01     !
!                         !                             !                      !
!      0.96849595E 00     !       0.83849059E 00        !   0.78621408E-01     !
!                         !                             !                      !
!      0.96807295E 00     !       0.83699439E 00        !   0.77867508E-01     !
!                         !                             !                      !
!      0.96764310E 00     !       0.83547959E 00        !   0.77116330E-01     !
!                         !                             !                      !
!      0.96720625E 00     !       0.83394584E 00        !   0.76367892E-01     !
!                         !                             !                      !
!      0.96676225E 00     !       0.83239291E 00        !   0.75622212E-01     !
--------------------------------------------------------------------------------
```

Table 6c. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS STATE AND SURNAME

SURNAME DECILE 3

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.85273914E 00	0.57117042E 00	0.23055
0.50%	0.9900	0.85140254E 00	0.56868052E 00	0.22830
1.00%	0.9801	0.85004955E 00	0.56617476E 00	0.22606
1.50%	0.9702	0.84867990E 00	0.56365305E 00	0.22384
2.00%	0.9604	0.84729337E 00	0.56111534E 00	0.22162
2.50%	0.9506	0.84588970E 00	0.55856164E 00	0.21941
3.00%	0.9409	0.84446864E 00	0.55599182E 00	0.21722
3.50%	0.9312	0.84302991E 00	0.55340588E 00	0.21503
4.00%	0.9216	0.84157327E 00	0.55080377E 00	0.21285
4.50%	0.9120	0.84009839E 00	0.54818542E 00	0.21069
5.00%	0.9025	0.83860506E 00	0.54555084E 00	0.20853
5.50%	0.8930	0.83709299E 00	0.54289994E 00	0.20639
6.00%	0.8836	0.83556188E 00	0.54023272E 00	0.20425

```
CISION OF RETRIEVED RESULTS FOR DATA FIELDS
 STATE
 AND SURNAME
--------------------------------------------------------------------------------
                                                                               !
      SURNAME DECILE  3                                                        !
-------------------------------------------------------------------------------!
!                            FILE SIZE                                         !
!------------------------------------------------------------------------------!
!       1 MILLION         !           4 MILLION         !     220 MILLION      !
-------------------------------------------------------------------------------!
!                         !                             !                      !
!     0.85273914E 00      !        0.57117042E 00       !   0.23055283E-01     !
!     0.85140254E 00      !        0.56860052E 00       !   0.22830506E-01     !
!     0.85004955E 00      !        0.56617476E 00       !   0.22606753E-01     !
!     0.84867990E 00      !        0.56365305E 00       !   0.22384025E-01     !
!     0.84729337E 00      !        0.56111534E 00       !   0.22162325E-01     !
!     0.84588970E 00      !        0.55856164E 00       !   0.21941655E-01     !
!     0.84446864E 00      !        0.55599182E 00       !   0.21722017E-01     !
!     0.84302991E 00      !        0.55340588E 00       !   0.21503407E-01     !
!     0.84157327E 00      !        0.55080377E 00       !   0.21285836E-01     !
!     0.84009839E 00      !        0.54818542E 00       !   0.21069296E-01     !
!     0.83860506E 00      !        0.54555084E 00       !   0.20853793E-01     !
!     0.83709299E 00      !        0.54289994E 00       !   0.20639331E-01     !
!     0.83556188E 00      !        0.54023272E 00       !   0.20425903E-01     !
--------------------------------------------------------------------------------
```

Table 6d. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS STATE AND SURNAME

SURNAME DECILE 4

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.65614718E 00	0.31682971E 00	0.82866
0.50%	0.9900	0.65383859E 00	0.31465378E 00	0.82046
1.00%	0.9801	0.65151126E 00	0.31247511E 00	0.81230
1.50%	0.9702	0.64916508E 00	0.31029379E 00	0.00410
2.00%	0.9604	0.64679989E 00	0.30810985E 00	0.79610
2.50%	0.9506	0.64441555E 00	0.30592344E 00	0.78806
3.00%	0.9409	0.64201197E 00	0.30373460E 00	0.78006
3.50%	0.9312	0.63958898E 00	0.30154344E 00	0.77210
4.00%	0.9216	0.63714647E 00	0.29935002E 00	0.76418
4.50%	0.9120	0.63468425E 00	0.29715442E 00	0.75630
5.00%	0.9025	0.63220224E 00	0.29495675E 00	0.74346
5.50%	0.8930	0.62970031E 00	0.29275709E 00	0.74056
6.00%	0.8836	0.62717831E 00	0.29055554E 00	0.73290

```
ON OF RETRIEVED RESULTS FOR DATA FIELDS
ATE
D SURNAME
--------------------------------------------------------------------------------
      SURNAME DECILE  4                                                        !
--------------------------------------------------------------------------------
                            FILE SIZE                                          !
--------------------------------------------------------------------------------
     1 MILLION        !           4 MILLION          !       220 MILLION       !
--------------------------------------------------------------------------------
                      !                              !                         !
  0.65614718E 00      !       0.31682971E 00         !     0.82866327E-02      !
  0.65303859E 00      !       0.31465378E 00         !     0.82046453E-02      !
  0.65151126E 00      !       0.31247511E 00         !     0.81230547E-02      !
  0.64916508E 00      !       0.31029379E 00         !     0.80410628E-02      !
  0.64679989E 00      !       0.30810985E 00         !     0.79610678E-02      !
  0.64441555E 00      !       0.30592344E 00         !     0.78806723E-02      !
  0.64201197E 00      !       0.30373460E 00         !     0.78006753E-02      !
  0.639[illegible]98E 00   !       0.30154344E 00         !     0.77210756E-02      !
  0.63714647E 00      !       0.29935002E 00         !     0.76418754E-02      !
  0.63468425E 00      !       0.29715442E 00         !     0.75630741E-02      !
  0.63220224E 00      !       0.29495675E 00         !     0.74346720E-02      !
  0.62970031E 00      !       0.29275709E 00         !     0.74066698E-02      !
  0.62717831E 00      !       0.29055554E 00         !     0.73290662E-02      !
--------------------------------------------------------------------------------
```

Table 6e. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STATE
AND SURNAME

SURNAME DECILE 5

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.51443352E 00	0.20679607E 00	0.469384
0.50%	0.9900	0.51190246E 00	0.20515223E 00	0.464723
1.00%	0.9801	0.50935818E 00	0.20350990E 00	0.460086
1.50%	0.9702	0.50680068E 00	0.20186917E 00	0.455471
2.00%	0.9604	0.50422990E 00	0.20023006E 00	0.450879
2.50%	0.9506	0.50164592E 00	0.19859271E 00	0.446310
3.00%	0.9409	0.49904872E 00	0.19695710E 00	0.441765
3.50%	0.9312	0.49643826E 00	0.19532333E 00	0.437242
4.00%	0.9216	0.49381460E 00	0.19369145E 00	0.432742
4.50%	0.9120	0.49117770E 00	0.19206154E 00	0.428265
5.00%	0.9025	0.48852757E 00	0.19043366E 00	0.423811
5.50%	0.8930	0.48586429E 00	0.18880785E 00	0.419380
6.00%	0.8836	0.48318781E 00	0.18718420E 00	0.414972

```
IEVED RESULTS FOR DATA FIELDS

-------------------------------------------------------------------------
NAME DECILE  5
-------------------------------------------------------------------------
                     FILE SIZE
-------------------------------------------------------------------------
LION      !           4 MILLION             !       220 MILLION          !
-------------------------------------------------------------------------
          !                                 !                            !
352E 00   !        0.20679607E 00           !     0.46938431E-02         !
246E 00   !        0.20515223E 00           !     0.46472372E-02         !
818E 00   !        0.20350990E 00           !     0.46008610E-02         !
068E 00   !        0.20186917E 00           !     0.45547142E-02         !
990E 00   !        0.20023006E 00           !     0.45087962E-02         !
592E 00   !        0.19859271E 00           !     0.44631082E-02         !
872E 00   !        0.19695710E 00           !     0.44176505E-02         !
826E 00   !        0.19532333E 00           !     0.43724211E-02         !
1460E 00  !        0.19369145E 00           !     0.43274221E-02         !
770E 00   !        0.19206154E 00           !     0.42826529E-02         !
2757E 00  !        0.19043366E 00           !     0.42381137E-02         !
5429E 00  !        0.18880785E 00           !     0.41938041E-02         !
8781E 00  !        0.18718420E 00           !     0.41497244E-02         !
-------------------------------------------------------------------------
```

Table 6f. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STATE
AND SURNAME

SURNAME DECILE 6

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.43377625E 00	0.15919675E 00	0.3418
0.50%	0.9500	0.43129673E 00	0.15785690E 00	0.3384
1.00%	0.9801	0.42880832E 00	0.15651959E 00	0.3350
1.50%	0.9702	0.42631104E 00	0.15518482E 00	0.3316
2.00%	0.9604	0.42380409E 00	0.15385265E 00	0.3283
2.50%	0.9506	0.42128994E 00	0.15252312E 00	0.3249
3.00%	0.9409	0.41876631E 00	0.15119627E 00	0.3216
3.50%	0.9312	0.41623400E 00	0.14987215E 00	0.3183
4.00%	0.9216	0.41369310E 00	0.14855082E 00	0.3150
4.50%	0.9120	0.41114365E 00	0.14723231E 00	0.3118
5.00%	0.9025	0.40858567E 00	0.14591665E 00	0.3085
5.50%	0.8930	0.40601932E 00	0.14460390E 00	0.3053
6.00%	0.8836	0.40344457E 00	0.14329411E 00	0.3021

```
IGN OF RETRIEVED RESULTS FOR DATA FIELDS
TATE
ND SURNAME
-----------------------------------------------------------------------------------------
                                                                                         !
        SURNAME DECILE  6                                                                !
-----------------------------------------------------------------------------------------!
                                 FILE SIZE                                               !
-----------------------------------------------------------------------------------------!
      1 MILLION          !             4 MILLION            !        220 MILLION         !
-----------------------------------------------------------------------------------------!
                         !                                  !                            !
    0.43377625E 00       !          0.15919675E 00          !      0.34180088E-02        !
    0.43129673E 00       !          0.15785690E 00          !      0.33840284E-02        !
    0.42880832E 00       !          0.15651959E 00          !      0.33502162E-02        !
    0.42631104E 00       !          0.15518482E 00          !      0.33165718E-02        !
    0.42380489E 00       !          0.15385265E 00          !      0.32830957E-02        !
    0.42128994E 00       !          0.15252312E 00          !      0.32497877E-02        !
    0.41876631E 00       !          0.15119627E 00          !      0.32166481E-02        !
    0.41623400E 00       !          0.14987215E 00          !      0.31836763E-02        !
    0.41369310E 00       !          0.14855082E 00          !      0.31508730E-02        !
    0.41114365E 00       !          0.14723231E 00          !      0.31182380E-02        !
    0.40858567E 00       !          0.14591665E 00          !      0.30857713E-02        !
    0.40601932E 00       !          0.14460390E 00          !      0.30534732E-02        !
    0.40344457E 00       !          0.14329411E 00          !      0.30213429E-02        !
-----------------------------------------------------------------------------------------
```

Table 6g. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STATE
AND SURNAME

SURNAME DECILE 7

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.28491025E 00	0.90091078E-01	0.17934
0.50%	0.9900	0.28286407E 00	0.89271841E-01	0.17756
1.00%	0.9801	0.28081657E 00	0.88455243E-01	0.17578
1.50%	0.9702	0.27876794E 00	0.87641339E-01	0.17401
2.00%	0.9604	0.27671813E 00	0.86830102E-01	0.17225
2.50%	0.9506	0.27466732E 00	0.86021594E-01	0.17050
3.00%	0.9409	0.27261553E 00	0.85215791E-01	0.16876
3.50%	0.9312	0.27056208E 00	0.84412743E-01	0.16703
4.00%	0.9216	0.26850942E 00	0.83612456E-01	0.16530
4.50%	0.9120	0.26645524E 00	0.82814944E-01	0.16359
5.00%	0.9025	0.26440041E 00	0.82020235E-01	0.16188
5.50%	0.8930	0.26234509E 00	0.81228333E-01	0.16019
6.00%	0.8836	0.26028529E 00	0.80439230E-01	0.15850

```
ION OF RETRIEVED RESULTS FOR DATA FIELDS
TATE
ND SURNAME
------------------------------------------------------------------------------------
        SURNAME DECILE  7                                                           !
------------------------------------------------------------------------------------!
                                   FILE SIZE                                        !
------------------------------------------------------------------------------------!
     1 MILLION           !           4 MILLION            !     220 MILLION         !
------------------------------------------------------------------------------------!
                         !                                !                         !
   0.28491025E 00        !        0.90091078E-01          !    0.17934645E-02       !
   0.28286407E 00        !        0.89271841E-01          !    0.17758062E-02       !
   0.28081657E 00        !        0.88455243E-01          !    0.17578367E-02       !
   0.27876794E 00        !        0.87641339E-01          !    0.17401562E-02       !
   0.27671813E 00        !        0.86830102E-01          !    0.17225645E-02       !
   0.27466732E 00        !        0.86021594E-01          !    0.17050616E-02       !
   0.27261553E 00        !        0.85215791E-01          !    0.16876480E-02       !
   0.27056288E 00        !        0.84412743E-01          !    0.16703229E-02       !
   0.26850942E 00        !        0.83612456E-01          !    0.16530872E-02       !
   0.26645524E 00        !        0.82814944E-01          !    0.16359401E-02       !
   0.26440041E 00        !        0.82020235E-01          !    0.16188821E-02       !
   0.26234509E 00        !        0.81228333E-01          !    0.16019131E-02       !
   0.26028929E 00        !        0.80439230E-01          !    0.15850331E-02       !
------------------------------------------------------------------------------------
```

Table 6h. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STATE
AND SURNAME

SURNAME DECILE 8

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.94171453E-01	0.25292678E-01	0.4713
0.50%	0.9900	0.93318859E-01	0.25046640E-01	0.4666
1.00%	0.9801	0.92468949E-01	0.24801710E-01	0.4619
1.50%	0.9702	0.91621762E-01	0.24557896E-01	0.4573
2.00%	0.9604	0.90777298E-01	0.24315193E-01	0.4526
2.50%	0.9506	0.89935575E-01	0.24073607E-01	0.4480
3.00%	0.9409	0.89096650E-01	0.23833141E-01	0.4434
3.50%	0.9312	0.88260499E-01	0.23593792E-01	0.4389
4.00%	0.9216	0.87427161E-01	0.23355566E-01	0.434
4.50%	0.9120	0.86596646E-01	0.23118401E-01	0.4290
5.00%	0.9025	0.85768983E-01	0.22882484E-01	0.4253
5.50%	0.8930	0.84944187E-01	0.22647629E-01	0.4209
6.00%	0.8836	0.84122285E-01	0.22413907E-01	0.4164

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 STATE
 AND SURNAME
----------------------------------------------------------------------------------
          SURNAME DECILE  8
----------------------------------------------------------------------------------
                            FILE SIZE
!---------------------------------------------------------------------------------
!                          !                          !       220 MILLION        !
!---------------------------------------------------------------------------------
!        1 MILLION         !        4 MILLION         !                          !
----------------------------------------------------------------------------------
!                          !                          !                          !
!      0.94171453E-01      !      0.25292678E-01      !      0.47133170E-03      !
!      0.93318859E-01      !      0.25046640E-01      !      0.46663236E-03      !
!      0.92468949E-01      !      0.24801710E-01      !      0.46195647E-03      !
!      0.91621762E-01      !      0.24557896E-01      !      0.45730417E-03      !
!      0.90777290E-01      !      0.24315193E-01      !      0.45267536E-03      !
!      0.89935575E-01      !      0.24073607E-01      !      0.44807003E-03      !
!      0.89096650E-01      !      0.23833141E-01      !      0.44348828E-03      !
!      0.88260499E-01      !      0.23593792E-01      !      0.43892994E-03      !
!      0.87427161E-01      !      0.23355566E-01      !      0.43439520E-03      !
!      0.86596646E-01      !      0.23118461E-01      !      0.42988401E-03      !
!      0.85768983E-01      !      0.22882484E-01      !      0.42539626E-03      !
!      0.84944187E-01      !      0.22647623E-01      !      0.42093208E-03      !
!      0.84122285E-01      !      0.22413907E-01      !      0.41649140E-03      !
----------------------------------------------------------------------------------
```

Table 6i. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STATE
AND SURNAME

SURNAME DECILE 9				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.60084473E-01	0.15714562E-01	0.2901
0.50%	0.9900	0.59520434E-01	0.15560223E-01	0.2872
1.00%	0.9801	0.58958556E-01	0.15406612E-01	0.2843
1.50%	0.9702	0.58398848E-01	0.15253726E-01	0.2814
2.00%	0.9604	0.57841309E-01	0.15101565E-01	0.2786
2.50%	0.9506	0.57285964E-01	0.14950136E-01	0.2757
3.00%	0.9409	0.56732816E-01	0.14799433E-01	0.2729
3.50%	0.9312	0.56181872E-01	0.14649459E-01	0.2701
4.00%	0.9216	0.55633149E-01	0.14500219E-01	0.2673
4.50%	0.9120	0.55086639E-01	0.14351706E-01	0.2645
5.00%	0.9025	0.54542362E-01	0.14203926E-01	0.2618
5.50%	0.8930	0.54000325E-01	0.14956877E-01	0.2590
6.00%	0.8836	0.53460536E-01	0.13910562E-01	0.2563

a

```
SION OF RETRIEVED RESULTS FOR DATA FIELDS
STATE
AND SURNAME
------------------------------------------------------------------------------------
                                                                                   !
      . SURNAME DECILE  9                                                          !
-----------------------------------------------------------------------------------!
                              FILE SIZE                                            !
-----------------------------------------------------------------------------------
    1 MILLION            !          4 MILLION           !      220 MILLION          !
-----------------------------------------------------------------------------------!
                         !                              !                           !
    0.60084473E-01       !       0.15714562E-01         !     0.29010226E-03        !
                         !                              !                           !
    0.59520434E-01       !       0.15560223E-01         !     0.28720932E-03        !
                         !                              !                           !
    0.58958556E-01       !       0.15406612E-01         !     0.28433086E-03        !
                         !                              !                           !
    0.58398848E-01       !       0.15253726E-01         !     0.28146687E-03        !
                         !                              !                           !
    0.57841309E-01       !       0.15101565E-01         !     0.27861738E-03        !
                         !                              !                           !
    0.57285964E-01       !       0.14950136E-01         !     0.27578238E-03        !
                         !                              !                           !
    0.56732816E-01       !       0.14799433E-01         !     0.27296189E-03        !
                         !                              !                           !
    0.56181872E-01       !       0.14649459E-01         !     0.27015580E-03        !
                         !                              !                           !
    0.55633149E-01       !       0.14500219E-01         !     0.26736429E-03        !
                         !                              !                           !
    0.55086639E-01       !       0.14351706E-01         !     0.26458723E-03        !
                         !                              !                           !
    0.54542362E-01       !       0.14203926E-01         !     0.26182466E-03        !
                         !                              !                           !
    0.54000325E-01       !       0.14056877E-01         !     0.25907658E-03        !
                         !                              !                           !
    0.53460536E-01       !       0.13910562E-01         !     0.25634296E-03        !
------------------------------------------------------------------------------------
```

Table 6j. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STATE
AND SURNAME

SURNAME DECILE 10				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.14565223E-01	0.36805747E-02	0.67156
0.50%	0.9900	0.14422007E-01	0.36439931E-02	0.6618
1.00%	0.9801	0.14279467E-01	0.36075924E-02	0.65820
1.50%	0.9702	0.14137604E-01	0.35713725E-02	0.6515
2.00%	0.9604	0.13996418E-01	0.35353330E-02	0.6449
2.50%	0.9506	0.13855913E-01	0.34994753E-02	0.6384
3.00%	0.9409	0.13716088E-01	0.34637977E-02	0.6318
3.50%	0.9312	0.13576941E-01	0.34283011E-02	0.6253
4.00%	0.9216	0.13439475E-01	0.33929855E-02	0.6189
4.50%	0.9120	0.13300689E-01	0.33578506E-02	0.6124
5.00%	0.9025	0.13163585E-01	0.33228972E-02	0.6060
5.50%	0.8930	0.13027164E-01	0.32881245E-02	0.5997
6.00%	0.8836	0.12891426E-01	0.32535332E-02	0.5934

ION OF RETRIEVED RESULTS FOR DATA FIELDS
TATE
ND SURNAME

SURNAME DECILE 10

FILE SIZE		
1 MILLION	4 MILLION	220 MILLION
0.14565223E-01	0.36805747E-02	0.67156551E-04
0.14422007E-01	0.36439931E-02	0.66486708E-04
0.14279467E-01	0.36075924E-02	0.65820217E-04
0.14137604E-01	0.35713725E-02	0.65157096E-04
0.13996418E-01	0.35353330E-02	0.64497323E-04
0.13855913E-01	0.34994753E-02	0.63840905E-04
0.13716088E-01	0.34637977E-02	0.63187851E-04
0.13576941E-01	0.34283011E-02	0.62538145E-04
0.13439475E-01	0.33929855E-02	0.61891805E-04
0.13300689E-01	0.33578506E-02	0.61248819E-04
0.13163585E-01	0.33228972E-02	0.60609189E-04
0.13027164E-01	0.32881245E-02	0.59972916E-04
0.12891426E-01	0.32535372E-02	0.59340000E-04

Table 7a. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 1				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.97577
0.50%	0.9851	0.10000000E 01	0.10000000E 01	0.97540
1.00%	0.9703	0.10000000E 01	0.10000000E 01	0.97503
1.50%	0.9557	0.10000000E 01	0.10000000E 01	0.97465
2.00%	0.9412	0.10000000E 01	0.10000000E 01	0.97427
2.50%	0.9269	0.10000000E 01	0.10000000E 01	0.97388
3.00%	0.9127	0.10000000E 01	0.10000000E 01	0.97347
3.50%	0.8986	0.10000000E 01	0.10000000E 01	0.97306
4.00%	0.8847	0.10000000E 01	0.10000000E 01	0.97265
4.50%	0.8710	0.10000000E 01	0.10000000E 01	0.97222
5.00%	0.9574	0.10000000E 01	0.10000000E 01	0.97179
5.50%	0.8439	0.10000000E 01	0.10000000E 01	0.97134
6.00%	0.8306	0.10000000E 01	0.10000000E 01	0.97089

```
RECISION OF RETRIEVED RESULTS FOR DATA FIELDS
     SEX
     DATE OF BIRTH (MONTH AND DAY ONLY)
     AND SURNAME
-------------------------------------------------------------------------------------
           SURNAME DECILE  1
-------------------------------------------------------------------------------------
R  !                             FILE SIZE                                           !
   !-----------------------------------------------------------------------------------
   !      1 MILLION          !         4 MILLION           !      220 MILLION        !
-------------------------------------------------------------------------------------
   !                         !                             !                         !
   !     0.10000000E 01      !       0.10000000E 01        !     0.97577313E 00      !
   !     0.10000000E 01      !       0.10000000E 01        !     0.97540930E 00      !
   !     0.10000000E 01      !       0.10000000E 01        !     0.97503831E 00      !
   !     0.10000000E 01      !       0.10000000E 01        !     0.97465997E 00      !
   !     0.10000000E 01      !       0.10000000E 01        !     0.97427411E 00      !
   !     0.10000000E 01      !       0.10000000E 01        !     0.97388055E 00      !
   !     0.10000000E 01      !       0.10000000E 01        !     0.97347909E 00      !
   !     0.10000000E 01      !       0.10000000E 01        !     0.97306955E 00      !
   !     0.10000000E 01      !       0.10000000E 01        !     0.97265173E 00      !
   !     0.10000000E 01      !       0.10000000E 01        !     0.97222541E 00      !
   !     0.10000000E 01      !       0.10000000E 01        !     0.97179038E 00      !
   !     0.10000000E 01      !       0.10000000E 01        !     0.97134643E 00      !
   !     0.10000000E 01      !       0.10000000E 01        !     0.97089333E 00      !
-------------------------------------------------------------------------------------
```

Table 7b. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 2

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.99825498E 00	0.98952098E 00	0.60707
0.50%	0.9851	0.99822252E 00	0.98935794E 00	0.60348
1.00%	0.9703	0.99818535E 00	0.98919158E 00	0.59985
1.50%	0.9557	0.99815544E 00	0.98902183E 00	0.59620
2.00%	0.9412	0.99812077E 00	0.98884859E 00	0.59252
2.50%	0.9269	0.99808532E 00	0.98867177E 00	0.58880
3.00%	0.9127	0.99804007E 00	0.98849129E 00	0.58506
3.50%	0.8986	0.99801201E 00	0.98830705E 00	0.58129
4.00%	0.8847	0.99797410E 00	0.98811896E 00	0.57749
4.50%	0.8710	0.99793532E 00	0.98792691E 00	0.57366
5.00%	0.8574	0.99789566E 00	0.98773081E 00	0.56980
5.50%	0.8439	0.99785508E 00	0.98753053E 00	0.56592
6.00%	0.8306	0.99781356E 00	0.98732598E 00	0.56200

```
ECISION OF RETRIEVED RESULTS FOR DATA FIELDS
   SEX
   DATE OF BIRTH (MONTH AND DAY ONLY)
   AND SURNAME
-------------------------------------------------------------------------------------------------
                                                                                                !
            SURNAME DECILE  2                                                                   !
-------------------------------------------------------------------------------------------------
 !                                      FILE SIZE                                               !
 !----------------------------------------------------------------------------------------------!
 !          1 MILLION           !               4 MILLION              !       220 MILLION      !
-------------------------------------------------------------------------------------------------
 !                              !                                      !                        !
 !       0.99825498E 00         !           0.98952098E 00             !     0.60707787E 00     !
 !                              !                                      !                        !
 !       0.99822252E 00         !           0.98935794E 00             !     0.60348300E 00     !
 !                              !                                      !                        !
 !       0.99818935E 00         !           0.98919158E 00             !     0.59985968E 00     !
 !                              !                                      !                        !
 !       0.99815544E 00         !           0.98902183E 00             !     0.59620496E 00     !
 !                              !                                      !                        !
 !       0.99812077E 00         !           0.98884859E 00             !     0.59252199E 00     !
 !                              !                                      !                        !
 !       0.99808532E 00         !           0.98867177E 00             !     0.58880951E 00     !
 !                              !                                      !                        !
 !       0.99804907E 00         !           0.98849129E 00             !     0.58506777E 00     !
 !                              !                                      !                        !
 !       0.99801201E 00         !           0.98830705E 00             !     0.58129687E 00     !
 !                              !                                      !                        !
 !       0.99797410E 00         !           0.98811896E 00             !     0.57749673E 00     !
 !                              !                                      !                        !
 !       0.99793532E 00         !           0.98792691E 00             !     0.57366754E 00     !
 !                              !                                      !                        !
 !       0.99789556E 00         !           0.98773081E 00             !     0.56980933E 00     !
 !                              !                                      !                        !
 !       0.99785508E 00         !           0.98753053E 00             !     0.56592220E 00     !
 !                              !                                      !                        !
 !       0.99781356E 00         !           0.98732598E 00             !     0.56200628E 00     !
-------------------------------------------------------------------------------------------------
```

Table 7c. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 3

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.98976009E 00	0.95694523E 00	0.282523
0.50%	0.9851	0.98960058E 00	0.95631581E 00	0.279484
1.00%	0.9703	0.98943782E 00	0.95567440E 00	0.27645
1.50%	0.9557	0.98927173E 00	0.95502073E 00	0.273422
2.00%	0.9412	0.98910223E 00	0.95435451E 00	0.270399
2.50%	0.9269	0.98892923E 00	0.95367546E 00	0.267382
3.00%	0.9127	0.98875265E 00	0.95298326E 00	0.26437
3.50%	0.8986	0.98857238E 00	0.95227762E 00	0.26136
4.00%	0.8847	0.98838834E 00	0.95155822E 00	0.25836
4.50%	0.8710	0.98820041E 00	0.95082472E 00	0.25537
5.00%	0.8574	0.98800852E 00	0.95007681E 00	0.252394
5.50%	0.8439	0.98781255E 00	0.94931413E 00	0.24941
6.00%	0.8306	0.98761240E 00	0.94853632E 00	0.24645

```
SION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME
-------------------------------------------------------------------------------------------
        SURNAME DECILE  3                                                                 !
-------------------------------------------------------------------------------------------!
                                FILE SIZE                                                 !
-------------------------------------------------------------------------------------------!
       1 MILLION          !            4 MILLION            !        220 MILLION          !
-------------------------------------------------------------------------------------------!
                          !                                 !                             !
     0.98976009E 00       !         0.95694523E 00          !      0.28252311E 00         !
                          !                                 !                             !
     0.98960058E 00       !         0.95631581E 00          !      0.27948449E 00         !
                          !                                 !                             !
     0.98943782E 00       !         0.95567440E 00          !      0.27645077E 00         !
                          !                                 !                             !
     0.98927173E 00       !         0.95502073E 00          !      0.27342224E 00         !
                          !                                 !                             !
     0.98910223E 00       !         0.95435451E 00          !      0.27039926E 00         !
                          !                                 !                             !
     0.98892923E 00       !         0.95367546E 00          !      0.26738213E 00         !
                          !                                 !                             !
     0.98875265E 00       !         0.95298326E 00          !      0.26437108E 00         !
                          !                                 !                             !
     0.98857238E 00       !         0.95227762E 00          !      0.26136648E 00         !
                          !                                 !                             !
     0.98838834E 00       !         0.95155822E 00          !      0.25836859E 00         !
                          !                                 !                             !
     0.98820041E 00       !         0.95082472E 00          !      0.25537774E 00         !
                          !                                 !                             !
     0.98800852E 00       !         0.95007681E 00          !      0.25239419E 00         !
                          !                                 !                             !
     0.98781255E 00       !         0.94931413E 00          !      0.24941830E 00         !
                          !                                 !                             !
     0.98761240E 00       !         0.94853632E 00          !      0.24645027E 00         !
-------------------------------------------------------------------------------------------
```

Table 7d. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

		SURNAME DECILE 4		
		FILE SIZE		
ERROR RATE	RECALL FACTOR	1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.96956024E 00	0.88556972E 00	0.12236
0.50%	0.9851	0.96910751E 00	0.88403220E 00	0.12075
1.00%	0.9703	0.96864596E 00	0.88243890E 00	0.11916
1.50%	0.9557	0.96817539E 00	0.88087936E 00	0.11757
2.00%	0.9412	0.96769558E 00	0.87926304E 00	0.11600
2.50%	0.9269	0.96720630E 00	0.87761945E 00	0.11443
3.00%	0.9127	0.96670733E 00	0.87594806E 00	0.11288
3.50%	0.8986	0.96619843E 00	0.87424837E 00	0.11134
4.00%	0.8847	0.96567936E 00	0.87251983E 00	0.10980
4.50%	0.8710	0.96514987E 00	0.87076186E 00	0.10828
5.00%	0.8574	0.96460970E 00	0.86897394E 00	0.10677
5.50%	0.8439	0.96405860E 00	0.86715547E 00	0.10527
6.00%	0.8306	0.96349628E 00	0.86530590E 00	0.10378

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 SEX
 DATE OF BIRTH (MONTH AND DAY ONLY)
 AND SURNAME
------------------------------------------------------------------------------------
          SURNAME DECILE  4                                                        !
-----------------------------------------------------------------------------------!
!                             FILE SIZE                                            !
!----------------------------------------------------------------------------------!
!       1 MILLION         !          4 MILLION           !       220 MILLION       !
------------------------------------------------------------------------------------
!                         !                              !                         !
!     0.96956024E 00      !       0.88556972E 00         !     0.12236357E 00      !
!     0.96910751E 00      !       0.88403220E 00         !     0.12075775E 00      !
!     0.96864596E 00      !       0.88243890E 00         !     0.11916217E 00      !
!     0.96817539E 00      !       0.88087936E 00         !     0.11757691E 00      !
!     0.96769558E 00      !       0.87926304E 00         !     0.11600202E 00      !
!     0.96720630E 00      !       0.87761945E 00         !     0.11443751E 00      !
!     0.96670733E 00      !       0.87594806E 00         !     0.11288367E 00      !
!     0.96619843E 00      !       0.87424837E 00         !     0.11134031E 00      !
!     0.96567936E 00      !       0.87251983E 00         !     0.10980755E 00      !
!     0.96514987E 00      !       0.87076186E 00         !     0.10828548E 00      !
!     0.96460970E 00      !       0.86897394E 00         !     0.10677412E 00      !
!     0.96405860E 00      !       0.86715547E 00         !     0.10527355E 00      !
!     0.96349628E 00      !       0.86530590E 00         !     0.10378379E 00      !
------------------------------------------------------------------------------------
```

Table 7e. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 5

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.94647884E 00	0.81310268E 00	0.72949
0.50%	0.9851	0.94570640E 00	0.81030261E 00	0.71938
1.00%	0.9703	0.94491952E 00	0.80846917E 00	0.70936
1.50%	0.9557	0.94411788E 00	0.80610190E 00	0.69941
2.00%	0.9412	0.94330113E 00	0.80370025E 00	0.68954
2.50%	0.9269	0.94246894E 00	0.80126377E 00	0.67976
3.00%	0.9127	0.94162096E 00	0.79879192E 00	0.6706
3.50%	0.8986	0.94075682E 00	0.79628422E 00	0.66042
4.00%	0.8847	0.93987616E 00	0.79374017E 00	0.65087
4.50%	0.8710	0.93897859E 00	0.79115919E 00	0.64141
5.00%	0.8574	0.93806374E 00	0.78854086E 00	0.63202
5.50%	0.8439	0.93713118E 00	0.78588457E 00	0.6227
6.00%	0.8305	0.93618052E 00	0.78318993E 00	0.61348

```
SION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME
-----------------------------------------------------------------------------
                                                                            !
     SURNAME DECILE  5
----------------------------------------------------------------------------!
                              FILE SIZE                                      !
----------------------------------------------------------------------------!
     1 MILLION          !         4 MILLION          !      220 MILLION      !
----------------------------------------------------------------------------!
                        !                            !                       !
   0.94647884E 00       !      0.81310268E 00        !   0.72949273E-01      !
   0.94570640E 00       !      0.81030261E 00        !   0.71938794E-01      !
   0.94491952E 00       !      0.80846917E 00        !   0.70936250E-01      !
   0.94411788E 00       !      0.80610190E 00        !   0.69941636E-01      !
   0.94330113E 00       !      0.80370025E 00        !   0.68954966E-01      !
   0.94246894E 00       !      0.80126377E 00        !   0.67976262E-01      !
   0.94162096E 00       !      0.79879192E 00        !   0.67006512E-01      !
   0.94075682E 00       !      0.79628422E 00        !   0.66042725E-01      !
   0.93987616E 00       !      0.79374017E 00        !   0.65087900E-01      !
   0.93897859E 00       !      0.79115919E 00        !   0.64141060E-01      !
   0.93806374E 00       !      0.78854086E 00        !   0.63202192E-01      !
   0.93713118E 00       !      0.78588457E 00        !   0.62271296E-01      !
   0.93618052E 00       !      0.78318993E 00        !   0.61348373E-01      !
-----------------------------------------------------------------------------
```

Table 7f. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 6

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.92747005E 00	0.75958315E 00	0.54129
0.50%	0.9851	0.92644658E 00	0.75682875E 00	0.53363
1.00%	0.9703	0.92540462E 00	0.75403292E 00	0.52606
1.50%	0.9557	0.92434375E 00	0.75120120E 00	0.51854
2.00%	0.9412	0.92326356E 00	0.74833322E 00	0.51109
2.50%	0.9269	0.92216364E 00	0.74542852E 00	0.50370
3.00%	0.9127	0.92104357E 00	0.74248676E 00	0.49637
3.50%	0.8986	0.91990291E 00	0.73950753E 00	0.4891
4.00%	0.8847	0.91874122E 00	0.73649048E 00	0.4819
4.50%	0.8710	0.91755803E 00	0.73343513E 00	0.47477
5.00%	0.8574	0.91635288E 00	0.73034114E 00	0.46770
5.50%	0.8439	0.91512528E 00	0.72720810E 00	0.46069
6.00%	0.8306	0.91387475E 00	0.72403563E 00	0.45375

```
CISION OF RETRIEVED RESULTS FOR DATA FIELDS
  SEX
  DATE OF BIRTH (MONTH AND DAY ONLY)
  AND SURNAME
-------------------------------------------------------------------------------

          SURNAME DECILE  6
-------------------------------------------------------------------------------!
!                             FILE SIZE                                        !
!------------------------------------------------------------------------------!
!        1 MILLION         !          4 MILLION          !     220 MILLION     !
-------------------------------------------------------------------------------!
!                          !                             !                     !
!     0.92747005E 00       !      0.75958915E 00         !   0.54129901E-01    !
!     0.92644658E 00       !      0.75682875E 00         !   0.53365107E-01    !
!     0.92540462E 00       !      0.75403292E 00         !   0.52606743E-01    !
!     0.92434375E 00       !      0.75120120E 00         !   0.51854788E-01    !
!     0.92326356E 00       !      0.74933322E 00         !   0.51109262E-01    !
!     0.92216364E 00       !      0.74542852E 00         !   0.50370152E-01    !
!     0.92104357E 00       !      0.74248676E 00         !   0.49637440E-01    !
!     0.91990291E 00       !      0.73950753E 00         !   0.48911140E-01    !
!     0.91874122E 00       !      0.73649048E 00         !   0.48191221E-01    !
!     0.91755803E 00       !      0.73343513E 00         !   0.47477703E-01    !
!     0.91635288E 00       !      0.73034114E 00         !   0.46770559E-01    !
!     0.91512528E 00       !      0.72720810E 00         !   0.46069790E-01    !
!     0.91387475E 00       !      0.72403563E 00         !   0.45375372E-01    !
-------------------------------------------------------------------------------
```

Table 7g. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 7

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.86928908E 00	0.62295784E 00	0.2910
0.50%	0.9951	0.86756610E 00	0.61941685E 00	0.2868
1.00%	0.9703	0.86581508E 00	0.61584515E 00	0.2826
1.50%	0.9557	0.86403560E 00	0.61224283E 00	0.2785
2.00%	0.9412	0.86222706E 00	0.60860372E 00	0.2744
2.50%	0.9269	0.86038999E 00	0.60494592E 00	0.2703
3.00%	0.9127	0.85852082E 00	0.60125131E 00	0.2663
3.50%	0.8986	0.85662204E 00	0.59752602E 00	0.2623
4.00%	0.8847	0.85469209E 00	0.59377002E 00	0.2583
4.50%	0.8710	0.85273040E 00	0.58998323E 00	0.2544
5.00%	0.8574	0.85073541E 00	0.58616596E 00	0.2505
5.50%	0.8439	0.84870954E 00	0.58231799E 00	0.2467
6.00%	0.8306	0.84664921E 00	0.57843954E 00	0.2429

e 7g. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 7

LL FACTOR	FILE SIZE		
	1 MILLION	4 MILLION	220 MILLION
1.0000	0.86928908E 00	0.62295784E 00	0.29106477E-01
0.9851	0.86756610E 00	0.61941685E 00	0.28684514E-01
0.9703	0.86581508E 00	0.61584515E 00	0.28266412E-01
0.9557	0.86403560E 00	0.61224283E 00	0.27852154E-01
0.9412	0.86222706E 00	0.60860972E 00	0.27441735E-01
0.9269	0.86038899E 00	0.60494592E 00	0.27035139E-01
0.9127	0.85852082E 00	0.60125131E 00	0.26632357E-01
0.8986	0.85662204E 00	0.59752602E 00	0.26233381E-01
0.8847	0.85469209E 00	0.59377002E 00	0.25838193E-01
0.8710	0.85273040E 00	0.58998329E 00	0.25446787E-01
0.8574	0.85073641E 00	0.58616596E 00	0.25059149E-01
0.8439	0.84870954E 00	0.58231799E 00	0.24675266E-01
0.8306	0.84664921E 00	0.57843954E 00	0.24295124E-01

Table 7h. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 8

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.63441127E 00	0.30217183E 00	0.78069
0.50%	0.9851	0.63091365E 00	0.29900975E 00	0.76911
1.00%	0.9703	0.62738447E 00	0.29585092E 00	0.75766
1.50%	0.9557	0.62382376E 00	0.29269574E 00	0.74633
2.00%	0.9412	0.62023138E 00	0.28954444E 00	0.73510
2.50%	0.9269	0.61660729E 00	0.28639735E 00	0.72399
3.00%	0.9127	0.61295153E 00	0.28325484E 00	0.71299
3.50%	0.8986	0.60926401E 00	0.28011719E 00	0.70209
4.00%	0.8847	0.60554473E 00	0.27698476E 00	0.6913
4.50%	0.8710	0.60179366E 00	0.27385780E 00	0.68064
5.00%	0.8574	0.59801089E 00	0.27073674E 00	0.67008
5.50%	0.8439	0.59419629E 00	0.26762179E 00	0.65962
6.00%	0.8306	0.59035003E 00	0.26451339E 00	0.64927

```
CISION OF RETRIEVED RESULTS FOR DATA FIELDS
   SEX
   DATE OF BIRTH (MONTH AND DAY ONLY)
   AND SURNAME
------------------------------------------------------------------------------

         SURNAME DECILE  8
------------------------------------------------------------------------------
!                                                                            !
!                               FILE SIZE                                    !
!----------------------------------------------------------------------------!
!      1 MILLION          !         4 MILLION          !      220 MILLION    !
------------------------------------------------------------------------------
!                         !                            !                     !
!    0.63441127E 00       !      0.30217183E 00        !   0.78068032E-02    !
!    0.63091365E 00       !      0.29900975E 00        !   0.76911826E-02    !
!    0.62738447E 00       !      0.29585092E 00        !   0.75766905E-02    !
!    0.62382376E 00       !      0.29269574E 00        !   0.74633228E-02    !
!    0.62023138E 00       !      0.28954444E 00        !   0.73510751E-02    !
!    0.61660729E 00       !      0.28639735E 00        !   0.72399418E-02    !
!    0.61295153E 00       !      0.28325484E 00        !   0.71299175E-02    !
!    0.60926401E 00       !      0.28011719E 00        !   0.70209981E-02    !
!    0.60554473E 00       !      0.27698476E 00        !   0.69131774E-02    !
!    0.60179366E 00       !      0.27385780E 00        !   0.68064531E-02    !
!    0.59801089E 00       !      0.27073674E 00        !   0.67008163E-02    !
!    0.59419629E 00       !      0.26762179E 00        !   0.65962642E-02    !
!    0.59035003E 00       !      0.26451339E 00        !   0.64927905E-02    !
------------------------------------------------------------------------------
```

Table 7i. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 9

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.51621502E 00	0.21037227E 00	0.48186
0.50%	0.9851	0.51245661E 00	0.20788480E 00	0.47470
1.00%	0.9703	0.50867785E 00	0.20540675E 00	0.46761
1.50%	0.9557	0.50487899E 00	0.20293829E 00	0.46060
2.00%	0.9412	0.50106018E 00	0.20047960E 00	0.45365
2.50%	0.9269	0.49722170E 00	0.19803091E 00	0.44677
3.00%	0.9127	0.49336395E 00	0.19559239E 00	0.43996
3.50%	0.8986	0.48948702E 00	0.19316424E 00	0.43323
4.00%	0.8847	0.48559127E 00	0.19074669E 00	0.42655
4.50%	0.8710	0.48167690E 00	0.18833984E 00	0.41995
5.00%	0.8574	0.47774432E 00	0.18594397E 00	0.41342
5.50%	0.8439	0.47379370E 00	0.18355921E 00	0.40695
6.00%	0.8306	0.46982546E 00	0.18118577E 00	0.40055

ble 7i. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 9

ECALL FACTOR	FILE SIZE		
	1 MILLION	4 MILLION	220 MILLION
1.0000	0.51621502E 00	0.21037227E 00	0.48186451E-02
0.9851	0.51245661E 00	0.20788430E 00	0.47470682E-02
0.9703	0.50867785E 00	0.20540675E 00	0.46761964E-02
0.9557	0.50487099E 00	0.20293829E 00	0.46060261E-02
0.9412	0.50106018E 00	0.20047960E 00	0.45365554E-02
0.9269	0.49722178E 00	0.19803091E 00	0.44677809E-02
0.9127	0.49336395E 00	0.19559239E 00	0.43996981E-02
0.8986	0.48948702E 00	0.19316424E 00	0.43323046E-02
0.8847	0.48559127E 00	0.19074669E 00	0.42655969E-02
0.8710	0.48167690E 00	0.18833984E 00	0.41995721E-02
0.8574	0.47774432E 00	0.18594397E 00	0.41342263E-02
0.8439	0.47379370E 00	0.18355921E 00	0.40695564E-02
0.8306	0.46982546E 00	0.18118577E 00	0.40055584E-02

Table 7j. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (MONTH AND DAY ONLY)
AND SURNAME

SURNAME DECILE 10

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.19789152E 00	0.58066060E-01	0.1119
0.50%	0.9851	0.19551480E 00	0.57248979E-01	0.1102
1.00%	0.9703	0.19314808E 00	0.56430674E-01	0.1086
1.50%	0.9557	0.19079135E 00	0.55635146E-01	0.1069
2.00%	0.9412	0.18844480E 00	0.54830382E-01	0.1053
2.50%	0.9269	0.18610867E 00	0.54048390E-01	0.1037
3.00%	0.9127	0.18378311E 00	0.53265149E-01	0.1021
3.50%	0.8986	0.18146829E 00	0.52488669E-01	0.1006
4.00%	0.8847	0.17916437E 00	0.51718943E-01	0.990
4.50%	0.8710	0.17687152E 00	0.50955950E-01	0.975
5.00%	0.8574	0.17458995E 00	0.50199711E-01	0.959
5.50%	0.8439	0.17231977E 00	0.49450193E-01	0.944
6.00%	0.8306	0.17006119E 00	0.48707404E-01	0.929

```
ION OF RETRIEVED RESULTS FOR DATA FIELDS
EX
ATE OF BIRTH (MONTH AND DAY ONLY)
ND SURNAME
----------------------------------------------------------------------------------
                                                                                 !
     SURNAME DECILE 10                                                           !
---------------------------------------------------------------------------------!
                         FILE SIZE                                               !
---------------------------------------------------------------------------------!
                    !                            !        220 MILLION            !
     1 MILLION      !        4 MILLION           !-------------------------------!
---------------------------------------------------------------------------------!
                    !                            !                               !
  0.19789152E 00    !     0.58066060E-01         !     0.11193779E-02            !
  0.19551400E 00    !     0.57248979E-01         !     0.11026896E-02            !
  0.19314008E 00    !     0.56438674E-01         !     0.10861674E-02            !
  0.19079135E 00    !     0.55635146E-01         !     0.10698109E-02            !
  0.18844400E 00    !     0.54838382E-01         !     0.10536189E-02            !
  0.18610867E 00    !     0.54048390E-01         !     0.10375909E-02            !
  0.18378311E 00    !     0.53265149E-01         !     0.10217258E-02            !
  0.18146829E 00    !     0.52488669E-01         !     0.10060231E-02            !
  0.17916437E 00    !     0.51718943E-01         !     0.99048161E-03            !
! 0.17687152E 00    !     0.50955950E-01         !     0.97510091E-03            !
! 0.17458995E 00    !     0.50199711E-01         !     0.95987991E-03            !
! 0.17231977E 00    !     0.49450193E-01         !     0.94481781E-03            !
! 0.17006119E 00    !     0.48707404E-01         !     0.92991377E-03            !
----------------------------------------------------------------------------------
```

Table 8a. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 1

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.99942
0.50%	0.9851	0.10000000E 01	0.10000000E 01	0.99941
1.00%	0.9703	0.10000000E 01	0.10000000E 01	0.99940
1.[illegible]0%	0.9557	0.10000000E 01	0.10000000E 01	0.99939
2.00%	0.9412	0.10000000E 01	0.10000000E 01	0.99938
2.50%	0.9269	0.10000000E 01	0.10000000E 01	0.99937
3.00%	0.9127	0.10000000E 01	0.10000000E 01	0.99936
3.50%	0.8986	0.10000000E 01	0.10000000E 01	0.99935
4.00%	0.8847	0.10000000E 01	0.10000000E 01	0.99934
4.50%	0.8710	0.10000000E 01	0.10000000E 01	0.99933
5.00%	0.8574	0.10000000E 01	0.10000000E 01	0.99932
5.50%	0.8439	0.10000000E 01	0.10000000E 01	0.99931
6.00%	0.8306	0.10000000E 01	0.10000000E 01	0.99930

```
ECISION OF RETRIEVED RESULTS FOR DATA FIELDS
   SEX
   DATE OF BIRTH (YEAR, MONTH, AND DAY)
   AND SURNAME
-----------------------------------------------------------------------------
                                                                            !
        SURNAME DECILE  1                                                   !
----------------------------------------------------------------------------!
 !                               FILE SIZE                                  !
 !--------------------------------------------------------------------------!
 !      1 MILLION          !          4 MILLION          !    220 MILLION   !
----------------------------------------------------------------------------!
 !                         !                             !                  !
 !    0.10000000E 01       !       0.10000000E 01        !  0.99942499E 00  !
 !                         !                             !                  !
 !    0.10000000E 01       !       0.10000000E 01        !  0.99941614E 00  !
 !                         !                             !                  !
 !    0.10000000E 01       !       0.10000000E 01        !  0.99940711E 00  !
 !                         !                             !                  !
 !    0.10000000E 01       !       0.10000000E 01        !  0.99939790E 00  !
 !                         !                             !                  !
 !    0.10000000E 01       !       0.10000000E 01        !  0.99938849E 00  !
 !                         !                             !                  !
 !    0.10000000E 01       !       0.10000000E 01        !  0.99937889E 00  !
 !                         !                             !                  !
 !    0.10000000E 01       !       0.10000000E 01        !  0.99936909E 00  !
 !                         !                             !                  !
 !    0.10000000E 01       !       0.10000000E 01        !  0.99935909E 00  !
 !                         !                             !                  !
 !    0.10000000E 01       !       0.10000000E 01        !  0.99934887E 00  !
 !                         !                             !                  !
 !    0.10000000E 01       !       0.10000000E 01        !  0.99933844E 00  !
 !                         !                             !                  !
 !    0.10000000E 01       !       0.10000000E 01        !  0.99932778E 00  !
 !                         !                             !                  !
 !    0.10000000E 01       !       0.10000000E 01        !  0.99931690E 00  !
 !                         !                             !                  !
 !    0.10000000E 01       !       0.10000000E 01        !  0.99930578E 00  !
-----------------------------------------------------------------------------
```

Table 8b. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 2				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MILL
0.00%	1.0000	0.99996020E 00	0.99975572E 00	0.985223
0.50%	0.9851	0.99995946E 00	0.99975107E 00	0.985002
1.00%	0.9703	0.99995870E 00	0.99974795E 00	0.984777
1.50%	0.9557	0.99995792E 00	0.99974394E 00	0.984548
2.00%	0.9412	0.99995712E 00	0.99973985E 00	0.984314
2.50%	0.9269	0.99995631E 00	0.99973567E 00	0.984075
3.00%	0.9127	0.99995547E 00	0.99973141E 00	0.983831
3.50%	0.8986	0.99995462E 00	0.99972705E 00	0.983582
4.00%	0.8847	0.99995375E 00	0.99972260E 00	0.983329
4.50%	0.8710	0.99995286E 00	0.99971806E 00	0.983070
5.00%	0.8574	0.99995195E 00	0.99971342E 00	0.982805
5.50%	0.8439	0.99995102E 00	0.99970868E 00	0.982536
6.00%	0.8306	0.99995006E 00	0.99970383E 00	0.982260

ECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 2

FILE SIZE		
1 MILLION	4 MILLION	220 MILLION
0.99996020E 00	0.99975572E 00	0.98522342E 00
0.99995946E 00	0.99975187E 00	0.98500275E 00
0.99995870E 00	0.99974795E 00	0.98477770E 00
0.99995792E 00	0.99974394E 00	0.98454816E 00
0.99995712E 00	0.99973985E 00	0.98431403E 00
0.99995631E 00	0.99973567E 00	0.98407518E 00
0.99995547E 00	0.99973141E 00	0.98383149E 00
0.99995462E 00	0.99972705E 00	0.98358286E 00
0.99995375E 00	0.99972260E 00	0.98332914E 00
0.99995286E 00	0.99971806E 00	0.98307021E 00
0.99995195E 00	0.99971342E 00	0.98280595E 00
0.99995102E 00	0.99970868E 00	0.98253621E 00
0.99995006E 00	0.99970383E 00	0.98226085E 00

Table 8c. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 3

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.99976451E 00	0.99896299E 00	0.944422
0.50%	0.9851	0.99976079E 00	0.99894713E 00	0.943627
1.00%	0.9703	0.99975698E 00	0.99893095E 00	0.942816
1.50%	0.9557	0.99975310E 00	0.99891444E 00	0.941993
2.00%	0.9412	0.99974913E 00	0.99889758E 00	0.941153
2.50%	0.9269	0.99974508E 00	0.99888038E 00	0.940298
3.00%	0.9127	0.99974095E 00	0.99886282E 00	0.939426
3.50%	0.8986	0.99973672E 00	0.99884489E 00	0.938537
4.00%	0.8847	0.99973241E 00	0.99882659E 00	0.937632
4.50%	0.8710	0.99972801E 00	0.99880790E 00	0.93670
5.00%	0.8574	0.99972351E 00	0.99878881E 00	0.93576
5.50%	0.8439	0.99971891E 00	0.99876932E 00	0.93481
6.00%	0.8306	0.99971421E 00	0.99874940E 00	0.93383

SION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 3

FILE SIZE		
1 MILLION	4 MILLION	220 MILLION
0.99976451E 00	0.99896299E 00	0.94442268E 00
0.99976079E 00	0.99894713E 00	0.94362794E 00
0.99975690E 00	0.99893095E 00	0.94281844E 00
0.99975310E 00	0.99891444E 00	0.94199384E 00
0.99974913E 00	0.99889758E 00	0.94115382E 00
0.99974508E 00	0.99888038E 00	0.94029801E 00
0.99974095E 00	0.99886282E 00	0.93942608E 00
0.99973672E 00	0.99884489E 00	0.93853764E 00
0.99973241E 00	0.99882659E 00	0.93763232E 00
0.99972801E 00	0.99880790E 00	0.93670974E 00
0.99972351E 00	0.99878881E 00	0.93576952E 00
0.99971891E 00	0.99876932E 00	0.93481124E 00
0.99971421E 00	0.99874940E 00	0.93387446E 00

Table 8d. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 4

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.99928573E 00	0.99702749E 00	0.857483
0.50%	0.9851	0.99927471E 00	0.99698238E 00	0.855635
1.00%	0.9703	0.99926347E 00	0.99693635E 00	0.853759
1.50%	0.9557	0.99925199E 00	0.99688940E 00	0.851851
2.00%	0.9412	0.99924028E 00	0.99684147E 00	0.849914
2.50%	0.9269	0.99922832E 00	0.99679257E 00	0.847946
3.00%	0.9127	0.99921612E 00	0.99674265E 00	0.845946
3.50%	0.8986	0.99920365E 00	0.99669170E 00	0.843915
4.00%	0.8847	0.99919093E 00	0.99663968E 00	0.841851
4.50%	0.8710	0.99917793E 00	0.99658657E 00	0.839754
5.00%	0.8574	0.99916466E 00	0.99653234E 00	0.837623
5.50%	0.8439	0.99915110E 00	0.99647696E 00	0.835458
6.00%	0.8306	0.99913725E 00	0.99642040E 00	0.833258

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 SEX
 DATE OF BIRTH (YEAR, MONTH, AND DAY)
 AND SURNAME
------------------------------------------------------------------------------
                                                                             !
         SURNAME DECILE  4                                                   !
-----------------------------------------------------------------------------!
!                            FILE SIZE                                       !
!----------------------------------------------------------------------------!
!        1 MILLION         !          4 MILLION          !    220 MILLION    !
-----------------------------------------------------------------------------!
!                          !                             !                   !
!     0.99928573E 00       !       0.99702749E 00        !  0.85740341E 00   !
!     0.99927471E 00       !       0.99698238E 00        !  0.85563568E 00   !
!     0.99926347E 00       !       0.99693635E 00        !  0.85375962E 00   !
!     0.99925199E 00       !       0.99688940E 00        !  0.85185172E 00   !
!     0.99924028E 00       !       0.99684147E 00        !  0.84991446E 00   !
!     0.99922832E 00       !       0.99679257E 00        !  0.84794634E 00   !
!     0.99921612E 00       !       0.99674265E 00        !  0.84594680E 00   !
!     0.99920365E 00       !       0.99669170E 00        !  0.84391530E 00   !
!     0.99919093E 00       !       0.99663966E 00        !  0.84185127E 00   !
!     0.99917793E 00       !       0.99658657E 00        !  0.83975420E 00   !
!     0.99916466E 00       !       0.99653234E 00        !  0.83762252E 00   !
!     0.99915110E 00       !       0.99647696E 00        !  0.83545863E 00   !
!     0.99913725E 00       !       0.99642040E 00        !  0.83325893E 00   !
------------------------------------------------------------------------------
```

Table 8e. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 5

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.99871424E 00	0.99472456E 00	0.772509
0.50%	0.9851	0.99869453E 00	0.99464400E 00	0.769855
1.00%	0.9703	0.99867441E 00	0.99455343E 00	0.767166
1.50%	0.9557	0.99865388E 00	0.99448041E 00	0.764442
2.00%	0.9412	0.99863293E 00	0.99439569E 00	0.761682
2.50%	0.9269	0.99861155E 00	0.99430925E 00	0.758885
3.00%	0.9127	0.99858972E 00	0.99422103E 00	0.756051
3.50%	0.8986	0.99856743E 00	0.99413030E 00	0.753100
4.00%	0.8847	0.99854468E 00	0.39403906E 00	0.750271
4.50%	0.8710	0.99852144E 00	0.99394522E 00	0.747324
5.00%	0.8574	0.99849771E 00	0.99384940E 00	0.744339
5.50%	0.8439	0.99847348E 00	0.99375157E 00	0.741315
6.00%	0.8306	0.99844872E 00	0.99365166E 00	0.738251

```
SION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME
-------------------------------------------------------------------------------------------
        SURNAME DECILE  5                                                                  !
-------------------------------------------------------------------------------------------!
                                    FILE SIZE                                              !
-------------------------------------------------------------------------------------------!
      1 MILLION            !            4 MILLION              !        220 MILLION        !
-------------------------------------------------------------------------------------------!
                           !                                   !                           !
    0.99871424E 00         !         0.99472456E 00            !      0.77250937E 00       !
    0.99869453E 00         !         0.99464430E 00            !      0.76985569E 00       !
    0.99867441E 00         !         0.99455343E 00            !      0.76716688E 00       !
    0.99865388E 00         !         0.99448041E 00            !      0.76444246E 00       !
    0.99863293E 00         !         0.99439569E 00            !      0.76168204E 00       !
    0.99861155E 00         !         0.99430925E 00            !      0.75888518E 00       !
    0.99858972E 00         !         0.99422103E 00            !      0.75605149E 00       !
    0.99856743E 00         !         0.99413098E 00            !      0.75318046E 00       !
    0.99854468E 00         !         0.99403906E 00            !      0.75027173E 00       !
    0.99852144E 00         !         0.99394522E 00            !      0.74732486E 00       !
    0.99849771E 00         !         0.99384940E 00            !      0.74433944E 00       !
    0.99847348E 00         !         0.99375157E 00            !      0.74131500E 00       !
    0.99844872E 00         !         0.99365166E 00            !      0.73825111E 00       !
-------------------------------------------------------------------------------------------
```

Table 8f. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 6

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MILL
0.00%	1.0000	0.99822275E 00	0.99275040E 00	0.711783
0.50%	0.9851	0.99819556E 00	0.99264107E 00	0.708688
1.00%	0.9703	0.99816783E 00	0.99252953E 00	0.705558
1.50%	0.9557	0.99813953E 00	0.99241575E 00	0.702392
2.00%	0.9412	0.99811064E 00	0.99229965E 00	0.699191
2.50%	0.9269	0.99808116E 00	0.99218118E 00	0.695953
3.00%	0.9127	0.99805106E 00	0.99206028E 00	0.692679
3.50%	0.8986	0.99802034E 00	0.99193690E 00	0.689369
4.00%	0.8847	0.99798937E 00	0.99181095E 00	0.686022
4.50%	0.8710	0.99795594E 00	0.99168239E 00	0.682638
5.00%	0.8574	0.99792423E 00	0.99155113E 00	0.679216
5.50%	0.8439	0.99789082E 00	0.99141712E 00	0.675757
6.00%	0.8306	0.99785670E 00	0.99128027E 00	0.67226

```
RECISION OF RETRIEVED RESULTS FOR DATA FIELDS
    SEX
    DATE OF BIRTH (YEAR, MONTH, AND DAY)
    AND SURNAME
-------------------------------------------------------------------------------
          SURNAME DECILE  6
-------------------------------------------------------------------------------
                              FILE SIZE
-------------------------------------------------------------------------------
R !                                                    !      220 MILLION     !
  !----------------------------------------------------------------------------!
  !      1 MILLION        !        4 MILLION           !                      !
-------------------------------------------------------------------------------
  !                       !                            !  0.71178324E 00      !
  !   0.99822275E 00      !    0.99275049E 00          !  0.70868833E 00      !
  !   0.99819556E 00      !    0.99264107E 00          !  0.70555826E 00      !
  !   0.99816783E 00      !    0.99252953E 00          !  0.70239261E 00      !
  !   0.99813953E 00      !    0.99241575E 00          !  0.69919123E 00      !
  !   0.99811064E 00      !    0.99229965E 00          !  0.69595377E 00      !
  !   0.99808116E 00      !    0.99218118E 00          !  0.69267999E 00      !
  !   0.99805196E 00      !    0.99206028E 00          !  0.68936964E 00      !
  !   0.99802034E 00      !    0.99193690E 00          !  0.68602239E 00      !
  !   0.99798937E 00      !    0.99181095E 00          !  0.68263809E 00      !
  !   0.99795594E 00      !    0.99168239E 00          !  0.67921646E 00      !
  !   0.99792423E 00      !    0.99155113E 00          !  0.67575727E 00      !
  !   0.99789082E 00      !    0.99141712E 00          !  0.67226020E 00      !
  !   0.99785670E 00      !    0.99128027E 00          !                      !
-------------------------------------------------------------------------------
```

Table 8g. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 7

ERROR RATE %	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MILL
0.00%	1.0000	0.99650832E 00	0.98622768E 00	0.5640272
0.50%	0.9851	0.99653636E 00	0.98602149E 00	0.5603258
1.00%	0.9703	0.99648334E 00	0.98581118E 00	0.5565990
1.50%	0.9557	0.99642923E 00	0.98559668E 00	0.5528469
2.00%	0.9412	0.99637402E 00	0.98537786E 00	0.5490696
2.50%	0.9269	0.99631768E 00	0.98515464E 00	0.5452673
3.00%	0.9127	0.99626017E 00	0.98492688E 00	0.5414401
3.50%	0.8986	0.99620146E 00	0.98469448E 00	0.5375982
4.00%	0.8847	0.99614152E 00	0.98445732E 00	0.5337118
4.50%	0.8710	0.99608032E 00	0.98421529E 00	0.5298110
5.00%	0.8574	0.99601783E 00	0.98396824E 00	0.5256861
5.50%	0.8439	0.99595402E 00	0.98371607E 00	0.5219372
6.00%	0.8306	0.99588884E 00	0.98345863E 00	0.5179646

```
RECISION OF RETRIEVED RESULTS FOR DATA FIELDS
    SEX
    DATE OF BIRTH (YEAR, MONTH, AND DAY)
    AND SURNAME
-----------------------------------------------------------------------------------
                                                                                  !
          SURNAME DECILE  7                                                       !
-----------------------------------------------------------------------------------
R  !                              FILE SIZE                                       !
   !------------------------------------------------------------------------------!
   !        1 MILLION         !          4 MILLION          !     220 MILLION     !
-----------------------------------------------------------------------------------
   !                          !                             !                     !
   !      0.99650832E 00      !       0.98622768E 00        !   0.56402726E 00    !
   !                          !                             !                     !
   !      0.99653636E 00      !       0.98602149E 00        !   0.56032585E 00    !
   !                          !                             !                     !
   !      0.99648334E 00      !       0.98581118E 00        !   0.55659905E 00    !
   !                          !                             !                     !
   !      0.99642923E 00      !       0.98559668E 00        !   0.55284693E 00    !
   !                          !                             !                     !
   !      0.99637402E 00      !       0.98537786E 00        !   0.54906967E 00    !
   !                          !                             !                     !
   !      0.99631768E 00      !       0.98515464E 00        !   0.54526733E 00    !
   !                          !                             !                     !
   !      0.99626017E 00      !       0.98492688E 00        !   0.54144017E 00    !
   !                          !                             !                     !
   !      0.99620146E 00      !       0.98469448E 00        !   0.53758827E 00    !
   !                          !                             !                     !
   !      0.99614152E 00      !       0.98445732E 00        !   0.53371183E 00    !
   !                          !                             !                     !
   !      0.99608032E 00      !       0.98421529E 00        !   0.52981103E 00    !
   !                          !                             !                     !
   !      0.99601783E 00      !       0.98396824E 00        !   0.52588614E 00    !
   !                          !                             !                     !
   !      0.99595402E 00      !       0.98371607E 00        !   0.52193728E 00    !
   !                          !                             !                     !
   !      0.99588884E 00      !       0.98345863E 00        !   0.51796469E 00    !
-----------------------------------------------------------------------------------
```

Table 8h. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 8				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.98705012E 00	0.94941197E 00	0.253478
0.50%	0.9851	0.98685517E 00	0.94868361E 00	0.250643
1.00%	0.9703	0.98665632E 00	0.94794159E 00	0.24781
1.50%	0.9557	0.98645349E 00	0.94718562E 00	0.24499
2.00%	0.9412	0.98624657E 00	0.94641536E 00	0.24218
2.50%	0.9269	0.98603546E 00	0.94563050E 00	0.23937
3.00%	0.9127	0.98582005E 00	0.94483070E 00	0.23657
3.50%	0.8986	0.98560023E 00	0.94401562E 00	0.23379
4.00%	0.8847	0.98537590E 00	0.94318492E 00	0.23101
4.50%	0.8710	0.98514694E 00	0.94233821E 00	0.22823
5.00%	0.8574	0.98491322E 00	0.94147514E 00	0.22547
5.50%	0.8439	0.98467463E 00	0.94059531E 00	0.22272
6.00%	0.8306	0.98443103E 00	0.93969836E 00	0.21998

8h. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 8

LL FACTOR	FILE SIZE		
	1 MILLION	4 MILLION	220 MILLION
1.0000	0.98705012E 00	0.94941197E 00	0.25347892E 00
0.9851	0.98685517E 00	0.94868361E 00	0.25064386E 00
0.9703	0.98665632E 00	0.94794159E 00	0.24781588E 00
0.9557	0.98645349E 00	0.94718562E 00	0.24499521E 00
0.9412	0.98624657E 00	0.94641536E 00	0.24218213E 00
0.9269	0.98603546E 00	0.94563050E 00	0.23937687E 00
0.9127	0.98582005E 00	0.94483070E 00	0.23657973E 00
0.8986	0.98560023E 00	0.94401562E 00	0.23379093E 00
0.8847	0.98537590E 00	0.94318492E 00	0.23101073E 00
0.8710	0.98514694E 00	0.94233821E 00	0.22823945E 00
0.8574	0.98491322E 00	0.94147514E 00	0.22547729E 00
0.8439	0.98467463E 00	0.94059531E 00	0.22272453E 00
0.8306	0.98443103E 00	0.93969836E 00	0.21998140E 00

Table 8i. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 9

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.97910909E 00	0.92029941E 00	0.172836
0.50%	0.9851	0.97879724E 00	0.91918763E 00	0.17069
1.00%	0.9703	0.97847924E 00	0.91805603E 00	0.16856
1.50%	0.9557	0.97815495E 00	0.91690419E 00	0.16645
2.00%	0.9412	0.97782420E 00	0.91573166E 00	0.16434
2.50%	0.9269	0.97748685E 00	0.91453805E 00	0.162246
3.00%	0.9127	0.97714272E 00	0.91332286E 00	0.160166
3.50%	0.8986	0.97679165E 00	0.91208567E 00	0.158086
4.00%	0.8847	0.97643346E 00	0.91082603E 00	0.156022
4.50%	0.8710	0.97606798E 00	0.90954341E 00	0.15397
5.00%	0.8574	0.97569503E 00	0.90823735E 00	0.151930
5.50%	0.8439	0.97531440E 00	0.[illegible]690735E 00	0.149902
6.00%	0.8306	0.97492592E 00	0.90555288E 00	0.147885

```
CISION OF RETRIEVED RESULTS FOR DATA FIELDS
 SEX
 DATE OF BIRTH (YEAR, MONTH, AND DAY)
 AND SURNAME
----------------------------------------------------------------------------------------
          SURNAME DECILE 9                                                              !
----------------------------------------------------------------------------------------!
!                              FILE SIZE                                                !
!---------------------------------------------------------------------------------------!
!        1 MILLION        !          4 MILLION            !         220 MILLION         !
----------------------------------------------------------------------------------------!
!                         !                               !                             !
!     0.97910909E 00      !       0.92029941E 00          !      0.17283696E 00         !
!     0.97879724E 00      !       0.91918763E 00          !      0.17069762E 00         !
!     0.97847924E 00      !       0.91805603E 00          !      0.16856875E 00         !
!     0.97815495E 00      !       0.91690419E 00          !      0.16645047E 00         !
!     0.97782420E 00      !       0.91573166E 00          !      0.16434295E 00         !
!     0.97748685E 00      !       0.91453805E 00          !      0.16224629E 00         !
!     0.97714272E 00      !       0.91332286E 00          !      0.16016065E 00         !
!     0.97679165E 00      !       0.91208567E 00          !      0.15808614E 00         !
!     0.97643346E 00      !       0.91082603E 00          !      0.15602290E 00         !
!     0.97606798E 00      !       0.90954341E 00          !      0.15397107E 00         !
!     0.97569503E 00      !       0.90823735E 00          !      0.15193078E 00         !
!     0.97531440E 00      !       0.90690735E 00          !      0.14990214E 00         !
!     0.97492592E 00      !       0.90555288E 00          !      0.14782525E 00         !
----------------------------------------------------------------------------------------
```

Table 8j. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 10

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.91551565E 00	0.72765376E 00	0.461291
0.50%	0.9851	0.91433796E 00	0.72465895E 00	0.454719
1.00%	0.9703	0.91313939E 00	0.72162055E 00	0.448204
1.50%	0.9557	0.91191953E 00	0.71856231E 00	0.441746
2.00%	0.9412	0.91067780E 00	0.71545984E 00	0.435344
2.50%	0.9269	0.90941412E 00	0.71232093E 00	0.428998
3.00%	0.9127	0.90812766E 00	0.70914512E 00	0.422710
3.50%	0.8986	0.90681806E 00	0.70593221E 00	0.416477
4.00%	0.8847	0.90548485E 00	0.70269188E 00	0.410301
4.50%	0.8710	0.90412753E 00	0.69939378E 00	0.404181
5.00%	0.8574	0.90274559E 00	0.69606765E 00	0.398117
5.50%	0.8439	0.90133851E 00	0.69270320E 00	0.392109
6.00%	0.8306	0.89990577E 00	0.68930013E 00	0.386157

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 SEX
 DATE OF BIRTH (YEAR, MONTH, AND DAY)
 AND SURNAME
----------------------------------------------------------------------------

          SURNAME DECILE 10
----------------------------------------------------------------------------
!                             FILE SIZE                                    !
!--------------------------------------------------------------------------!
!      1 MILLION         !          4 MILLION          !     220 MILLION     !
----------------------------------------------------------------------------
!                        !                             !                     !
!    0.91551565E 00      !       0.72765376E 00        !   0.46129155E-01    !
!                        !                             !                     !
!    0.91433796E 00      !       0.72465895E 00        !   0.45471950E-01    !
!                        !                             !                     !
!    0.91313939E 00      !       0.72162855E 00        !   0.44820443E-01    !
!                        !                             !                     !
!    0.91191953E 00      !       0.71856231E 00        !   0.44174602E-01    !
!                        !                             !                     !
!    0.91067790E 00      !       0.71545984E 00        !   0.43534425E-01    !
!                        !                             !                     !
!    0.90941412E 00      !       0.71232093E 00        !   0.42899896E-01    !
!                        !                             !                     !
!    0.90812766E 00      !       0.70914512E 00        !   0.42271018E-01    !
!                        !                             !                     !
!    0.90681806E 00      !       0.70593221E 00        !   0.41647774E-01    !
!                        !                             !                     !
!    0.90548485E 00      !       0.70268188E 00        !   0.41030152E-01    !
!                        !                             !                     !
!    0.90412753E 00      !       0.69939378E 00        !   0.40418151E-01    !
!                        !                             !                     !
!    0.90274559E 00      !       0.69606765E 00        !   0.39811760E-01    !
!                        !                             !                     !
!    0.90133851E 00      !       0.69270320E 00        !   0.39210960E-01    !
!                        !                             !                     !
!    0.89990577E 00      !       0.68930013E 00        !   0.38615741E-01    !
----------------------------------------------------------------------------
```

Table 9a. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS

STREET NAME
STREET NUMBER
STATE
AND SURNAME

SURNAME DECILE 1

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.99999
0.50%	0.9801	0.10000000E 01	0.10000000E 01	0.99999
1.00%	0.9606	0.10000000E 01	0.10000000E 01	0.99999
1.50%	0.9413	0.10000000E 01	0.10000000E 01	0.99999
2.00%	0.9224	0.10000000E 01	0.10000000E 01	0.99999
2.50%	0.9037	0.10000000E 01	0.10000000E 01	0.99999
3.00%	0.8853	0.10000000E 01	0.10000000E 01	0.99999
3.50%	0.8672	0.10000000E 01	0.10000000E 01	0.99999
4.00%	0.8493	0.10000000E 01	0.10000000E 01	0.99999
4.50%	0.8318	0.10000000E 01	0.10000000E 01	0.99999
5.00%	0.8145	0.10000000E 01	0.10000000E 01	0.99999
5.50%	0.7975	0.10000000E 01	0.10000000E 01	0.99999
6.00%	0.7807	0.10000000E 01	0.10000000E 01	0.99999

```
ION OF RETRIEVED RESULTS FOR DATA FIELDS
TREET NAME
TREET NUMBER
TATE
ND SURNAME
--------------------------------------------------------------------------------------
     SURNAME DECILE  1                                                               !
-------------------------------------------------------------------------------------!
                              FILE SIZE                                              !
-------------------------------------------------------------------------------------!
      1 MILLION         !          4 MILLION           !        220 MILLION          !
-------------------------------------------------------------------------------------!
                        !                              !                             !
   0.10000000E 01       !       0.10000000E 01         !     0.99999999E 00          !
   0.10000000E 01       !       0.10000000E 01         !     0.99999999E 00          !
   0.10000000E 01       !       0.10000000E 01         !     0.99999999E 00          !
   0.10000000E 01       !       0.10000000E 01         !     0.99999999E 00          !
   0.10000000E 01       !       0.10000000E 01         !     0.99999999E 00          !
   0.10000000E 01       !       0.10000000E 01         !     0.99999999E 00          !
   0.10000000E 01       !       0.10000000E 01         !     0.99999999E 00          !
   0.10000000E 01       !       0.10000000E 01         !     0.99999999E 00          !
   0.10000000E 01       !       0.10000000E 01         !     0.99999999E 00          !
   0.10000000E 01       !       0.10000000E 01         !     0.99999999E 00          !
   0.10000000E 01       !       0.10000000E 01         !     0.99999999E 00          !
   0.10000000E 01       !       0.10000000E 01         !     0.99999999E 00          !
   0.10000000E 01       !       0.10000000E 01         !     0.99999999E 00          !
--------------------------------------------------------------------------------------
```

Table 9b. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NAME
STREET NUMBER
STATE
AND SURNAME

SURNAME DECILE 2				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MILL
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.999999
0.50%	0.9801	0.10000000E 01	0.10000000E 01	0.999999
1.00%	0.9606	0.10000000E 01	0.10000000E 01	0.999999
1.50%	0.9413	0.10000000E 01	0.10000000E 01	0.999999
2.00%	0.9224	0.10000000E 01	0.10000000E 01	0.999999
2.50%	0.9037	0.10000000E 01	0.10000000E 01	0.999999
3.00%	0.8853	0.10000000E 01	0.10000000E 01	0.999999
3.50%	0.8672	0.10000000E 01	0.10000000E 01	0.999999
4.00%	0.8493	0.10000000E 01	0.10000000E 01	0.999999
4.50%	0.8318	0.10000000E 01	0.10000000E 01	0.999999
5.00%	0.8145	0.10000000E 01	0.10000000E 01	0.999999
5.50%	0.7975	0.10000000E 01	0.10000000E 01	0.999999
6.00%	0.7807	0.10000000E 01	0.10000000E 01	0.999999

```
ION OF RETRIEVED RESULTS FOR DATA FIELDS
TREET NAME
TREET NUMBER
TATE
ND SURNAME
-------------------------------------------------------------------------------------------
                                                                                          !
       SURNAME DECILE  2                                                                  !
------------------------------------------------------------------------------------------!
                               FILE SIZE                                                  !
------------------------------------------------------------------------------------------!
      1 MILLION          !          4 MILLION            !        220 MILLION             !
------------------------------------------------------------------------------------------!
                         !                               !                                !
   0.10000000E 01        !       0.10000000E 01          !      0.99999978E 00            !
                         !                               !                                !
   0.10000000E 01        !       0.10000000E 01          !      0.99999978E 00            !
                         !                               !                                !
   0.10000000E 01        !       0.10000000E 01          !      0.99999978E 00            !
                         !                               !                                !
   0.10000000E 01        !       0.10000000E 01          !      0.99999977E 00            !
                         !                               !                                !
   0.10000000E 01        !       0.10000000E 01          !      0.99999977E 00            !
                         !                               !                                !
   0.10000000E 01        !       0.10000000E 01          !      0.99999976E 00            !
                         !                               !                                !
   0.10000000E 01        !       0.10000000E 01          !      0.99999976E 00            !
                         !                               !                                !
   0.10000000E 01        !       0.10000000E 01          !      0.99999975E 00            !
                         !                               !                                !
   0.10000000E 01        !       0.10000000E 01          !      0.99999975E 00            !
                         !                               !                                !
   0.10000000E 01        !       0.10000000E 01          !      0.99999974E 00            !
                         !                               !                                !
   0.10000000E 01        !       0.10000000E 01          !      0.99999973E 00            !
                         !                               !                                !
   0.10000000E 01        !       0.10000000E 01          !      0.99999973E 00            !
                         !                               !                                !
   0.10000000E 01        !       0.10000000E 01          !      0.99999972E 00            !
-------------------------------------------------------------------------------------------
```

Table 9c. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NAME
STREET NUMBER
STATE
AND SURNAME

SURNAME DECILE 3

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	280 MIL
0.00%	1.0000	0.10000000E 01	0.99999999E 00	0.999999
0.50%	0.9801	0.10000000E 01	0.99999998E 00	0.999999
1.00%	0.9606	0.10000000E 01	0.99999998E 00	0.999999
1.50%	0.9413	0.10000000E 01	0.99999998E 00	0.999999
2.00%	0.9224	0.10000000E 01	0.99999998E 00	0.999999
2.50%	0.9037	0.10000000E 01	0.99999998E 00	0.999999
3.00%	0.8853	0.10000000E 01	0.99999998E 00	0.999999
3.50%	0.8672	0.10000000E 01	0.99999998E 00	0.999999
4.00%	0.8493	0.10000000E 01	0.99999998E 00	0.999999
4.50%	0.8318	0.10000000E 01	0.99999998E 00	0.999999
5.00%	0.8145	0.10000000E 01	0.99999998E 00	0.999998
5.50%	0.7975	0.10000000E 01	0.99999998E 00	0.999998
6.00%	0.7807	0.10000000E 01	0.99999998E 00	0.999998

```
ION OF RETRIEVED RESULTS FOR DATA FIELDS
TREET NAME
TREET NUMBER
TATE
ND SURNAME
--------------------------------------------------------------------------------------
                                                                                     !
     SURNAME DECILE  3                                                               !
--------------------------------------------------------------------------------------!
                              FILE SIZE                                              !
--------------------------------------------------------------------------------------
      1 MILLION          !          4 MILLION          !        220 MILLION          !
--------------------------------------------------------------------------------------!
                         !                             !                             !
    0.10000000E 01       !       0.99999999E 00        !       0.99999915E 00        !
    0.10000000E 01       !       0.99999998E 00        !       0.99999914E 00        !
    0.10000000E 01       !       0.99999998E 00        !       0.99999912E 00        !
    0.10000000E 01       !       0.99999998E 00        !       0.99999910E 00        !
    0.10000000E 01       !       0.99999998E 00        !       0.99999908E 00        !
    0.10000000E 01       !       0.99999998E 00        !       0.99999906E 00        !
    0.10000000E 01       !       0.99999998E 00        !       0.99999904E 00        !
    0.10000000E 01       !       0.99999998E 00        !       0.99999902E 00        !
    0.10000000E 01       !       0.99999998E 00        !       0.99999900E 00        !
    0.10000000E 01       !       0.99999998E 00        !       0.99999898E 00        !
    0.10000000E 01       !       0.99999998E 00        !       0.99999896E 00        !
    0.10000000E 01       !       0.99999998E 00        !       0.99999894E 00        !
    0.10000000E 01       !       0.99999998E 00        !       0.99999891E 00        !
--------------------------------------------------------------------------------------
```

Table 9d. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NAME
STREET NUMBER
STATE
AND SURNAME

SURNAME DECILE 4

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.99999999E 00	0.99999996E 00	0.99999
0.50%	0.9901	0.99999999E 00	0.99999996E 00	0.99999
1.00%	0.9606	0.99999999E 00	0.99999996E 00	0.99999
1.50%	0.9413	0.99999999E 00	0.99999995E 00	0.99999
2.00%	0.9224	0.99999999E 00	0.99999995E 00	0.99999
2.50%	0.9037	0.99999999E 00	0.99999995E 00	0.99999
3.00%	0.8853	0.99999999E 00	0.99999995E 00	0.99999
3.50%	0.8672	0.99999999E 00	0.99999995E 00	0.99999
4.00%	0.8493	0.99999999E 00	0.99999995E 00	0.99999
4.50%	0.8318	0.99999999E 00	0.99999995E 00	0.99999
5.00%	0.8145	0.99999999E 00	0.99999995E 00	0.99999
5.50%	0.7975	0.99999999E 00	0.99999995E 00	0.99999
6.00%	0.7837	0.99999999E 00	0.99999994E 00	0.99999

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 STREET NAME
 STREET NUMBER
 STATE
 AND SURNAME
-----------------------------------------------------------------------------!
                                                                             !
          SURNAME DECILE  4                                                  !
-----------------------------------------------------------------------------!
!                              FILE SIZE                                     !
!----------------------------------------------------------------------------!
!       1 MILLION            !        4 MILLION          !    220 MILLION    !
-----------------------------------------------------------------------------!
!                            !                           !                   !
!     0.99999999E 00         !     0.99999996E 00        !  0.99999761E 00   !
!     0.99999999E 00         !     0.99999996E 00        !  0.99999756E 00   !
!     0.99999999E 00         !     0.99999996E 00        !  0.99999751E 00   !
!     0.99999999E 00         !     0.99999995E 00        !  0.99999746E 00   !
!     0.99999999E 00         !     0.99999995E 00        !  0.99999741E 00   !
!     0.99999999E 00         !     0.99999995E 00        !  0.99999735E 00   !
!     0.99999999E 00         !     0.99999995E 00        !  0.99999730E 00   !
!     0.99999999E 00         !     0.99999995E 00        !  0.99999724E 00   !
!     0.99999999E 00         !     0.99999995E 00        !  0.99999718E 00   !
!     0.99999999E 00         !     0.99999995E 00        !  0.99999712E 00   !
!     0.99999999E 00         !     0.99999995E 00        !  0.99999706E 00   !
!     0.99999999E 00         !     0.99999995E 00        !  0.99999700E 00   !
!     0.99999993E 00         !     0.99999994E 00        !  0.99999693E 00   !
-----------------------------------------------------------------------------
```

Table 9e. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NAME
STREET NUMBER
STATE
AND SURNAME

SURNAME DECILE 5

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.99999998E 00	0.99999992E 00	0.999995
0.50%	0.9801	0.99999998E 00	0.99999992E 00	0.999995
1.00%	0.9606	0.99999998E 00	0.99999992E 00	0.999995
1.50%	0.9413	0.99999998E 00	0.99999992E 00	0.999995
2.00%	0.9224	0.99999998E 00	0.99999992E 00	0.999995
2.50%	0.9037	0.99999998E 00	0.99999992E 00	0.999995
3.00%	0.8853	0.99999998E 00	0.99999991E 00	0.999995
3.50%	0.8672	0.99999998E 00	0.99999991E 00	0.999995
4.00%	0.8493	0.99999998E 00	0.99999991E 00	0.999995
4.50%	0.8318	0.99999998E 00	0.99999991E 00	0.999994
5.00%	0.8145	0.99999998E 00	0.99999991E 00	0.999994
5.50%	0.7975	0.99999998E 00	0.99999990E 00	0.999994
6.00%	0.7807	0.99999998E 00	0.99999990E 00	0.999994

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
  STREET NAME
  STREET NUMBER
  STATE
  AND SURNAME
----------------------------------------------------------------------------------
            SURNAME DECILE  5
----------------------------------------------------------------------------------
!                                FILE SIZE                                        !
!---------------------------------------------------------------------------------!
!        1 MILLION         !          4 MILLION          !        220 MILLION       !
-----------------------------------------------------------------------------------
!                          !                             !                          !
!     0.99999998E 00       !       0.99999992E 00        !     0.99999576E 00       !
!     0.99999998E 00       !       0.99999992E 00        !     0.99999567E 00       !
!     0.99999998E 00       !       0.99999992E 00        !     0.99999559E 00       !
!     0.99999998E 00       !       0.99999992E 00        !     0.99999550E 00       !
!     0.99999998E 00       !       0.99999992E 00        !     0.99999540E 00       !
!     0.99999998E 00       !       0.99999992E 00        !     0.99999531E 00       !
!     0.99999998E 00       !       0.99999991E 00        !     0.99999521E 00       !
!     0.99999998E 00       !       0.99999991E 00        !     0.99999511E 00       !
!     0.99999998E 00       !       0.99999991E 00        !     0.99999501E 00       !
!     0.99999998E 00       !       0.99999991E 00        !     0.99999490E 00       !
!     0.99999998E 00       !       0.99999991E 00        !     0.99999479E 00       !
!     0.99999998E 00       !       0.99999990E 00        !     0.99999468E 00       !
!     0.99999998E 00       !       0.99999990E 00        !     0.99999457E 00       !
-----------------------------------------------------------------------------------
```

Table 9f. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NAME
STREET NUMBER
STATE
AND SURNAME

SURNAME DECILE 6

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.99999997E 00	0.99999989E 00	0.999994
0.50%	0.9801	0.99999997E 00	0.99999989E 00	0.999994
1.00%	0.9606	0.99999997E 00	0.99999989E 00	0.999993
1.50%	0.9413	0.99999997E 00	0.99999989E 00	0.999993
2.00%	0.9224	0.99999997E 00	0.99999989E 00	0.999993
2.50%	0.9037	0.99999997E 00	0.99999988E 00	0.999993
3.00%	0.8853	0.99999997E 00	0.99999988E 00	0.999993
3.50%	0.8672	0.99999997E 00	0.99999988E 00	0.999993
4.00%	0.8493	0.99999997E 00	0.99999988E 00	0.999993
4.50%	0.8318	0.99999997E 00	0.99999987E 00	0.999992
5.00%	0.8145	0.99999997E 00	0.99999987E 00	0.999992
5.50%	0.7975	0.99999997E 00	0.99999987E 00	0.999992
6.00%	0.7807	0.99999997E 00	0.99999987E 00	0.999992

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 STREET NAME
 STREET NUMBER
 STATE
 AND SURNAME
----------------------------------------------------------------------------------
                                                                                  !
        SURNAME DECILE  6                                                         !
----------------------------------------------------------------------------------!
!                           FILE SIZE                                             !
!---------------------------------------------------------------------------------!
!        1 MILLION          !         4 MILLION          !      220 MILLION       !
----------------------------------------------------------------------------------!
!                           !                            !                        !
!      0.99999997E 00       !     0.99999989E 00         !   0.99999417E 00       !
!                           !                            !                        !
!      0.99999997E 00       !     0.99999989E 00         !   0.99999405E 00       !
!                           !                            !                        !
!      0.99999997E 00       !     0.99999989E 00         !   0.99999393E 00       !
!                           !                            !                        !
!      0.99999997E 00       !     0.99999989E 00         !   0.99999381E 00       !
!                           !                            !                        !
!      0.99999997E 00       !     0.99999989E 00         !   0.99999368E 00       !
!                           !                            !                        !
!      0.99999997E 00       !     0.99999988E 00         !   0.99999355E 00       !
!                           !                            !                        !
!      0.99999997E 00       !     0.99999988E 00         !   0.99999341E 00       !
!                           !                            !                        !
!      0.99999997E 00       !     0.99999988E 00         !   0.99999328E 00       !
!                           !                            !                        !
!      0.99999997E 00       !     0.99999988E 00         !   0.99999313E 00       !
!                           !                            !                        !
!      0.99999997E 00       !     0.99999987E 00         !   0.99999299E 00       !
!                           !                            !                        !
!      0.99999997E 00       !     0.99999987E 00         !   0.99999284E 00       !
!                           !                            !                        !
!      0.99999997E 00       !     0.99999987E 00         !   0.99999269E 00       !
!                           !                            !                        !
!      0.99999997E 00       !     0.99999987E 00         !   0.99999253E 00       !
----------------------------------------------------------------------------------
```

Table 9g. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NAME
STREET NUMBER
STATE
AND SURNAME

SURNAME DECILE 7

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.99999995E 00	0.99999980E 00	0.999988
0.50%	0.9801	0.99999995E 00	0.99999979E 00	0.999988
1.00%	0.9606	0.99999995E 00	0.99999979E 00	0.999988
1.50%	0.9413	0.99999995E 00	0.99999979E 00	0.999988
2.00%	0.9224	0.99999995E 00	0.99999978E 00	0.999987
2.50%	0.9037	0.99999995E 00	0.99999978E 00	0.999987
3.00%	0.8853	0.99999994E 00	0.99999977E 00	0.999987
3.50%	0.8672	0.99999994E 00	0.99999977E 00	0.999987
4.00%	0.8493	0.99999994E 00	0.99999976E 00	0.999986
4.50%	0.8318	0.99999994E 00	0.99999976E 00	0.999986
5.00%	0.8145	0.99999994E 00	0.99999975E 00	0.999986
5.50%	0.7975	0.99999994E 00	0.99999975E 00	0.999986
6.00%	0.7807	0.99999994E 00	0.99999974E 00	0.999985

```
CISION OF RETRIEVED RESULTS FOR DATA FIELDS
   STREET NAME
   STREET NUMBER
   STATE
   AND SURNAME
------------------------------------------------------------------------------

        SURNAME DECILE  7                                                    !
------------------------------------------------------------------------------
 !                          FILE SIZE                                        !
 !----------------------------------------------------------------------------!
 !      1 MILLION         !         4 MILLION         !      220 MILLION     !
------------------------------------------------------------------------------
 !                        !                           !                      !
 !    0.99999995E 00      !      0.99999980E 00       !    0.99998887E 00    !
 !    0.99999995E 00      !      0.99999979E 00       !    0.99998864E 00    !
 !    0.99999995E 00      !      0.99999979E 00       !    0.99998841E 00    !
 !    0.99999995E 00      !      0.99999979E 00       !    0.99998818E 00    !
 !    0.99999995E 00      !      0.99999978E 00       !    0.99998793E 00    !
 !    0.99999995E 00      !      0.99999978E 00       !    0.99998768E 00    !
 !    0.99999994E 00      !      0.99999977E 00       !    0.99998743E 00    !
 !    0.99999994E 00      !      0.99999977E 00       !    0.99998716E 00    !
 !    0.99999994E 00      !      0.99999976E 00       !    0.99998689E 00    !
 !    0.99999994E 00      !      0.99999976E 00       !    0.99998662E 00    !
 !    0.99999994E 00      !      0.99999975E 00       !    0.99998633E 00    !
 !    0.99999994E 00      !      0.99999975E 00       !    0.99998604E 00    !
 !    0.99999994E 00      !      0.99999974E 00       !    0.99998574E 00    !
------------------------------------------------------------------------------
```

Table 9h. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NAME
STREET NUMBER
STATE
AND SURNAME

SURNAME DECILE 8

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.99999981E 00	0.99999923E 00	0.99995
0.50%	0.9801	0.99999981E 00	0.99999922E 00	0.99995
1.00%	0.9606	0.99999980E 00	0.99999920E 00	0.99995
1.50%	0.9413	0.99999980E 00	0.99999918E 00	0.99995
2.00%	0.9224	0.99999979E 00	0.99999917E 00	0.99995
2.50%	0.9037	0.99999979E 00	0.99999915E 00	0.99995
3.00%	0.8853	0.99999979E 00	0.99999913E 00	0.99995
3.50%	0.8672	0.99999978E 00	0.99999911E 00	0.99995
4.00%	0.8493	0.99999978E 00	0.99999910E 00	0.99995
4.50%	0.8318	0.99999977E 00	0.99999908E 00	0.99994
5.00%	0.8145	0.99999977E 00	0.99999906E 00	0.99994
5.50%	0.7975	0.99999976E 00	0.99999904E 00	0.99994
6.00%	0.7807	0.99999976E 00	0.99999902E 00	0.99994

ble 9h. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NAME
STREET NUMBER
STATE
AND SURNAME

SURNAME DECILE 8

ECALL FACTOR	FILE SIZE		
	1 MILLION	4 MILLION	220 MILLION
1.0000	0.99999981E 00	0.99999923E 00	0.99995759E 00
0.9801	0.99999981E 00	0.99999922E 00	0.99995673E 00
0.9606	0.99999980E 00	0.99999920E 00	0.99995585E 00
0.9413	0.99999980E 00	0.99999918E 00	0.99995495E 00
0.9224	0.99999979E 00	0.99999917E 00	0.99995402E 00
0.9037	0.99999979E 00	0.99999915E 00	0.99995307E 00
0.8853	0.99999979E 00	0.99999913E 00	0.99995210E 00
0.8672	0.99999978E 00	0.99999911E 00	0.99995110E 00
0.8493	0.99999978E 00	0.99999910E 00	0.99995007E 00
0.8310	0.99999977E 00	0.99999908E 00	0.99994902E 00
0.8145	0.99999977E 00	0.99999906E 00	0.99994793E 00
0.7975	0.99999976E 00	0.99999904E 00	0.99994682E 00
0.7807	0.99999976E 00	0.99999902E 00	0.99994568E 00

Table 9i. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NAME
STREET NUMBER
STATE
AND SURNAME

SURNAME DECILE 9

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.99999969E 00	0.99999875E 00	0.99993
0.50%	0.9801	0.99999969E 00	0.99999873E 00	0.99992
1.00%	0.9606	0.99999968E 00	0.99999870E 00	0.99992
1.50%	0.9413	0.99999967E 00	0.99999867E 00	0.99992
2.00%	0.9224	0.99999967E 00	0.99999865E 00	0.99992
2.50%	0.9037	0.99999966E 00	0.99999862E 00	0.99992
3.00%	0.8853	0.99999965E 00	0.99999859E 00	0.99992
3.50%	0.8672	0.99999964E 00	0.99999856E 00	0.99992
4.00%	0.8493	0.99999964E 00	0.99999853E 00	0.9999
4.50%	0.8318	0.99999963E 00	0.99999850E 00	0.9999
5.00%	0.8145	0.99999962E 00	0.99999847E 00	0.9999
5.50%	0.7975	0.99999961E 00	0.99999843E 00	0.9999
6.00%	0.7807	0.99999960E 00	0.99999840E 00	0.9999

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 STREET NAME
 STREET NUMBER
 STATE
 AND SURNAME
------------------------------------------------------------------------------------------
                                                                                         !
        SURNAME DECILE  9                                                                !
-----------------------------------------------------------------------------------------!
!                                  FILE SIZE                                             !
!----------------------------------------------------------------------------------------!
!       1 MILLION           !           4 MILLION           !        220 MILLION         !
-----------------------------------------------------------------------------------------!
!                           !                               !                            !
!     0.99999969E 00        !        0.99999875E 00         !      0.99993109E 00        !
!                           !                               !                            !
!     0.99999969E 00        !        0.99999873E 00         !      0.99992969E 00        !
!                           !                               !                            !
!     0.99999968E 00        !        0.99999870E 00         !      0.99992826E 00        !
!                           !                               !                            !
!     0.99999967E 00        !        0.99999867E 00         !      0.99992679E 00        !
!                           !                               !                            !
!     0.99999967E 00        !        0.99999865E 00         !      0.99992529E 00        !
!                           !                               !                            !
!     0.99999966E 00        !        0.99999862E 00         !      0.99992374E 00        !
!                           !                               !                            !
!     0.99999965E 00        !        0.99999859E 00         !      0.99992216E 00        !
!                           !                               !                            !
!     0.99999964E 00        !        0.99999856E 00         !      0.99992053E 00        !
!                           !                               !                            !
!     0.99999964E 00        !        0.99999853E 00         !      0.99991887E 00        !
!                           !                               !                            !
!     0.99999963E 00        !        0.99999850E 00         !      0.99991715E 00        !
!                           !                               !                            !
!     0.99999962E 00        !        0.99999847E 00         !      0.99991540E 00        !
!                           !                               !                            !
!     0.99999961E 00        !        0.99999843E 00         !      0.99991359E 00        !
!                           !                               !                            !
!     0.99999960E 00        !        0.99999840E 00         !      0.99991174[illegible] !
------------------------------------------------------------------------------------------
```

Table 9j. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
STREET NAME
STREET NUMBER
STATE
AND SURNAME

SURNAME DECILE 10				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.99999867E 00	0.99999461E 00	0.99970
0.50%	0.9801	0.99999864E 00	0.99999450E 00	0.99969
1.00%	0.9606	0.99999861E 00	0.99999438E 00	0.99969
1.50%	0.9413	0.99999858E 00	0.99999427E 00	0.99968
2.00%	0.9224	0.99999855E 00	0.99999415E 00	0.99967
2.50%	0.9037	0.99999852E 00	0.99999403E 00	0.99967
3.00%	0.8853	0.99999849E 00	0.99999391E 00	0.99966
3.50%	0.8672	0.99999846E 00	0.99999378E 00	0.99965
4.00%	0.8493	0.99999843E 00	0.99999365E 00	0.99964
4.50%	0.8318	0.99999840E 00	0.99999351E 00	0.99964
5.00%	0.8145	0.99999836E 00	0.99999338E 00	0.99963
5.50%	0.7975	0.99999833E 00	0.99999324E 00	0.99962
6.00%	0.7807	0.99999829E 00	0.99999309E 00	0.9996

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 STREET NAME
 STREET NUMBER
 STATE
 AND SURNAME
-----------------------------------------------------------------------------------------------
            SURNAME DECILE 10
-----------------------------------------------------------------------------------------------
!                                   FILE SIZE                                                 !
!---------------------------------------------------------------------------------------------!
!       1 MILLION           !          4 MILLION              !        220 MILLION            !
-----------------------------------------------------------------------------------------------
!                           !                                 !                               !
!     0.95999857E 00        !       0.99999461E 00            !      0.99970232E 00           !
!     0.99999864E 00        !       0.99999450E 00            !      0.99969629E 00           !
!     0.99999861E 00        !       0.99999438E 00            !      0.99969011E 00           !
!     0.99999858E 00        !       0.99999427E 00            !      0.99968377E 00           !
!     0.99999855E 00        !       0.99999415E 00            !      0.99967727E 00           !
!     0.99999852E 00        !       0.99999403E 00            !      0.99967060E 00           !
!     0.99999849E 00        !       0.99999391E 00            !      0.99966376E 00           !
!     0.99999846E 00        !       0.99999378E 00            !      0.99965674E 00           !
!     0.99999843E 00        !       0.99999365E 00            !      0.99964953E 00           !
!     0.99999840E 00        !       0.99999351E 00            !      0.99964214E 00           !
!     0.99999836E 00        !       0.99999338E 00            !      0.99963455E 00           !
!     0.99999833E 00        !       0.99999324E 00            !      0.99962675E 00           !
!     0.99999829E 00        !       0.99999309E 00            !      0.99961875E 00           !
-----------------------------------------------------------------------------------------------
```

Table 10a. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
STREET NAME
STREET NUMBER
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE I

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.10000
0.50%	0.9752	0.10000000E 01	0.10000000E 01	0.10000
1.00%	0.9510	0.10000000E 01	0.10000000E 01	0.10000
1.50%	0.9272	0.10000000E 01	0.10000000E 01	0.10000
2.00%	0.9039	0.10000000E 01	0.10000000E 01	0.10000
2.50%	0.8811	0.10000000E 01	0.10000000E 01	0.10000
3.00%	0.8597	0.10000000E 01	0.10000000E 01	0.10000
3.50%	0.8368	0.10000000E 01	0.10000000E 01	0.10000
4.00%	0.8154	0.10000000E 01	0.10000000E 01	0.10000
4.50%	0.7944	0.10000000E 01	0.10000000E 01	0.10000
5.00%	0.7738	0.10000000E 01	0.10000000E 01	0.10000
5.50%	0.7536	0.10000000E 01	0.10000000E 01	0.10000
6.00%	0.7339	0.10000000E 01	0.10000000E 01	0.10000

```
CISION OF RETRIEVED RESULTS FOR DATA FIELDS
  SEX
  STREET NAME
  STREET NUMBER
  DATE OF BIRTH (YEAR, MONTH, AND DAY)
  AND SURNAME
-------------------------------------------------------------------------------
!                                                                             !
        SURNAME DECILE  1                                                     !
-------------------------------------------------------------------------------
!                              FILE SIZE                                      !
!-----------------------------------------------------------------------------!
!          1 MILLION          !          4 MILLION          !   220 MILLION   !
-------------------------------------------------------------------------------
!                             !                             !                 !
!      0.10000000E 01         !      0.10000000E 01         !  0.10000000E 01 !
!                             !                             !                 !
!      0.10000000E 01         !      0.10000000E 01         !  0.10000000E 01 !
!                             !                             !                 !
!      0.10000000E 01         !      0.10000000E 01         !  0.10000000E 01 !
!                             !                             !                 !
!      0.10000000E 01         !      0.10000000E 01         !  0.10000000E 01 !
!                             !                             !                 !
!      0.10000000E 01         !      0.10000000E 01         !  0.10000000E 01 !
!                             !                             !                 !
!      0.10000000E 01         !      0.10000000E 01         !  0.10000000E 01 !
!                             !                             !                 !
!      0.10000000E 01         !      0.10000000E 01         !  0.10000000E 01 !
!                             !                             !                 !
!      0.10000000E 01         !      0.10000000E 01         !  0.10000000E 01 !
!                             !                             !                 !
!      0.10000000E 01         !      0.10000000E 01         !  0.10000000E 01 !
!                             !                             !                 !
!      0.10000000E 01         !      0.10000000E 01         !  0.10000000E 01 !
!                             !                             !                 !
!      0.10000000E 01         !      0.10000000E 01         !  0.10000000E 01 !
!                             !                             !                 !
!      0.10000000E 01         !      0.10000000E 01         !  0.10000000E 01 !
!                             !                             !                 !
!      0.10000000E 01         !      0.10000000E 01         !  0.10000000E 01 !
-------------------------------------------------------------------------------
```

Table 10b. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
STREET NAME
STREET NUMBER
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 2

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MILL
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.1000000
0.50%	0.9752	0.10000000E 01	0.10000000E 01	0.1000000
1.00%	0.9510	0.10000000E 01	0.10000000E 01	0.1000000
1.50%	0.9272	0.10000000E 01	0.10000000E 01	0.1000000
2.00%	0.9039	0.10000000E 01	0.10000000E 01	0.1000000
2.50%	0.8811	0.10000000E 01	0.10000000E 01	0.1000000
3.00%	0.8587	0.10000000E 01	0.10000000E 01	0.1000000
3.50%	0.8368	0.10000000E 01	0.10000000E 01	0.1000000
4.00%	0.8154	0.10000000E 01	0.10000000E 01	0.1000000
4.50%	0.7944	0.10000000E 01	0.10000000E 01	0.1000000
5.00%	0.7738	0.10000000E 01	0.10000000E 01	0.1000000
5.50%	0.7536	0.10000000E 01	0.10000000E 01	0.1000000
6.00%	0.7339	0.10000000E 01	0.10000000E 01	0.1000000

```
ECISION OF RETRIEVED RESULTS FOR DATA FIELDS
    SEX
    STREET NAME
    STREET NUMBER
    DATE OF BIRTH (YEAR, MONTH, AND DAY)
    AND SURNAME
--------------------------------------------------------------------------------

          SURNAME DECILE  2
--------------------------------------------------------------------------------
  !                                 FILE SIZE                                  !
  !----------------------------------------------------------------------------!
  !        1 MILLION           !        4 MILLION          !     220 MILLION    !
--------------------------------------------------------------------------------
  !                            !                           !                    !
  !     0.10000000E 01         !     0.10000000E 01        !   0.10000000E 01   !
  !     0.10000000E 01         !     0.10000000E 01        !   0.10000000E 01   !
  !     0.10000000E 01         !     0.10000000E 01        !   0.10000000E 01   !
  !     0.10000000E 01         !     0.10000000E 01        !   0.10000000E 01   !
  !     0.10000000E 01         !     0.10000000E 01        !   0.10000000E 01   !
  !     0.10000000E 01         !     0.10000000E 01        !   0.10000000E 01   !
  !     0.10000000E 01         !     0.10000000E 01        !   0.10000000E 01   !
  !     0.10000000E 01         !     0.10000000E 01        !   0.10000000E 01   !
  !     0.10000000E 01         !     0.10000000E 01        !   0.10000000E 01   !
  !     0.10000000E 01         !     0.10000000E 01        !   0.10000000E 01   !
  !     0.10000000E 01         !     0.10000000E 01        !   0.10000000E 01   !
  !     0.10000000E 01         !     0.10000000E 01        !   0.10000000E 01   !
  !     0.10000000E 01         !     0.10000000E 01        !   0.10000000E 01   !
--------------------------------------------------------------------------------
```

Table 10c. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
STREET NAME
STREET NUMBER
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 3

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MIL
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.10000
0.50%	0.9752	0.10000000E 01	0.10000000E 01	0.10000
1.00%	0.9510	0.10000000E 01	0.10000000E 01	0.10000
1.50%	0.9272	0.10000000E 01	0.10000000E 01	0.10000
2.00%	0.9039	0.10000000E 01	0.10000000E 01	0.10000
2.50%	0.8811	0.10000000E 01	0.10000000E 01	0.10000
3.00%	0.8587	0.10000000E 01	0.10000000E 01	0.10000
3.50%	0.8368	0.10000000E 01	0.10000000E 01	0.10000
4.00%	0.8154	0.10000000E 01	0.10000000E 01	0.10000
4.50%	0.7944	0.10000000E 01	0.10000000E 01	0.10000
5.00%	0.7738	0.10000000E 01	0.10000000E 01	0.10000
5.50%	0.7536	0.10000000E 01	0.10000000E 01	0.10000
6.00%	0.7339	0.10000000E 01	0.10000000E 01	0.10000

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 SEX
 STREET NAME
 STREET NUMBER
 DATE OF BIRTH (YEAR, MONTH, AND DAY)
 AND SURNAME
------------------------------------------------------------------------------------
                                                                                   !
        SURNAME DECILE  3                                                          !
------------------------------------------------------------------------------------
!                                FILE SIZE                                         !
!-----------------------------------------------------------------------------------!
!          1 MILLION          !          4 MILLION          !     220 MILLION      !
-------------------------------------------------------------------------------------
!                             !                             !                      !
!       0.10000000E 01        !       0.10000000E 01        !   0.10000000E 01     !
!       0.10000000E 01        !       0.10000000E 01        !   0.10000000E 01     !
!       0.10000000E 01        !       0.10000000E 01        !   0.10000000E 01     !
!       0.10000000E 01        !       0.10000000E 01        !   0.10000000E 01     !
!       0.10000000E 01        !       0.10000000E 01        !   0.10000000E 01     !
!       0.10000000E 01        !       0.10000000E 01        !   0.10000000E 01     !
!       0.10000000E 01        !       0.10000000E 01        !   0.10000000E 01     !
!       0.10000000E 01        !       0.10000000E 01        !   0.10000000E 01     !
!       0.10000000E 01        !       0.10000000E 01        !   0.10000000E 01     !
!       0.10000000E 01        !       0.10000000E 01        !   0.10000000E 01     !
!       0.10000000E 01        !       0.10000000E 01        !   0.10000000E 01     !
!       0.10000000E 01        !       0.10000000E 01        !   0.10000000E 01     !
!       0.10000000E 01        !       0.10000000E 01        !   0.10000000E 01     !
-------------------------------------------------------------------------------------
```

Table 10d. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
STREET NAME
STREET NUMBER
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 4

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MI
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.10000
0.50%	0.9752	0.10000000E 01	0.10000000E 01	0.10000
1.00%	0.9510	0.10000000E 01	0.10000000E 01	0.10000
1.50%	0.9272	0.10000000E 01	0.10000000E 01	0.10000
2.00%	0.9039	0.10000000E 01	0.10000000E 01	0.10000
2.50%	0.8811	0.10000000E 01	0.10000000E 01	0.10000
3.00%	0.8587	0.10000000E 01	0.10000000E 01	0.10000
3.50%	0.8368	0.10000000E 01	0.10000000E 01	0.10000
4.00%	0.8154	0.10000000E 01	0.10000000E 01	0.10000
4.50%	0.7944	0.10000000E 01	0.10000000E 01	0.10000
5.00%	0.7738	0.10000000E 01	0.10000000E 01	0.10000
5.50%	0.7536	0.10000000E 01	0.10000000E 01	0.10000
6.00%	0.7339	0.10000000E 01	0.10000000E 01	0.10000

```
ECISION OF RETRIEVED RESULTS FOR DATA FIELDS
    SEX
    STREET NAME
    STREET NUMBER
    DATE OF BIRTH (YEAR, MONTH, AND DAY)
    AND SURNAME
-----------------------------------------------------------------------------

          SURNAME DECILE  4
-----------------------------------------------------------------------------
 !                                                                           !
 !                               FILE SIZE                                   !
 !---------------------------------------------------------------------------!
 !      1 MILLION          !         4 MILLION          !    220 MILLION *    !
-----------------------------------------------------------------------------
 !                         !                            !                    !
 !     0.10000000E 01      !      0.10000000E 01        !   0.10000000E 01   !
 !     0.10000000E 01      !      0.10000000E 01        !   0.10000000E 01   !
 !     0.10000000E 01      !      0.10000000E 01        !   0.10000000E 01   !
 !     0.10000000E 01      !      0.10000000E 01        !   0.10000000E 01   !
 !     0.10000000E 01      !      0.10000000E 01        !   0.10000000E 01   !
 !     0.10000000E 01      !      0.10000000E 01        !   0.10000000E 01   !
 !     0.10000000E 01      !      0.10000000E 01        !   0.10000000E 01   !
 !     0.10000000E 01      !      0.10000000E 01        !   0.10000000E 01   !
 !     0.10000000E 01      !      0.10000000E 01        !   0.10000000E 01   !
 !     0.10000000E 01      !      0.10000000E 01        !   0.10000000E 01   !
 !     0.10000000E 01      !      0.10000000E 01        !   0.10000000E 01   !
 !     0.10000000E 01      !      0.10000000E 01        !   0.10000000E 01   !
 !     0.10000000E 01      !      0.10000000E 01        !   0.10000000E 01   !
-----------------------------------------------------------------------------
```

Table 10e. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
STREET NAME
STREET NUMBER
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 5

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.9999
0.50%	0.9752	0.10000000E 01	0.10000000E 01	0.9999
1.00%	0.9510	0.10000000E 01	0.10000000E 01	0.9999
1.50%	0.9272	0.10000000E 01	0.10000000E 01	0.9999
2.00%	0.9039	0.10000000E 01	0.10000000E 01	0.9999
2.50%	0.8811	0.10000000E 01	0.10000000E 01	0.9999
3.00%	0.8587	0.10000000E 01	0.10000000E 01	0.9999
3.50%	0.8368	0.10000000E 01	0.10000000E 01	0.9999
4.90%	0.9154	0.10000000E 01	0.10000000E 01	0.9999
4.50%	0.7944	0.10000000E 01	0.10000000E 01	0.9999
5.00%	0.7738	0.10000000E 01	0.10000000E 01	0.9999
5.50%	0.7536	0.10000000E 01	0.10000000E 01	0.9999
6.00%	0.7339	0.10000000E 01	0.10000000E 01	0.9999

```
CISION OF RETRIEVED RESULTS FOR DATA FIELDS
  SEX
  STREET NAME
  STREET NUMBER
  DATE OF BIRTH (YEAR, MONTH, AND DAY)
  AND SURNAME
---------------------------------------------------------------------------
                                                                          !
        SURNAME DECILE  5                                                 !
--------------------------------------------------------------------------!
!                             FILE SIZE                                   !
!-------------------------------------------------------------------------!
!        1 MILLION        !         4 MILLION         !    220 MILLION    !
---------------------------------------------------------------------------
!                         !                           !                   !
!     0.10000000E 01      !      0.10000000E 01       !  0.99999999E 00   !
!     0.10000000E 01      !      0.10000000E 01       !  0.99999999E 00   !
!     0.10000000E 01      !      0.10000000E 01       !  0.99999999E 00   !
!     0.10000000E 01      !      0.10000000E 01       !  0.99999999E 00   !
!     0.10000000E 01      !      0.10000000E 01       !  0.99999999E 00   !
!     0.10000000E 01      !      0.10000000E 01       !  0.99999999E 00   !
!     0.10000000E 01      !      0.10000000E 01       !  0.99999999E 00   !
!     0.10000000E 01      !      0.10000000E 01       !  0.99999999E 00   !
!     0.10000000E 01      !      0.10000000E 01       !  0.99999999E 00   !
!     0.10000000E 01      !      0.10000000E 01       !  0.99999999E 00   !
!     0.10000000E 01      !      0.10000000E 01       !  0.99999999E 00   !
!     0.10000000E 01      !      0.10000000E 01       !  0.99999999E 00   !
!     0.10000000E 01      !      0.10000000E 01       !  0.99999999E 00   !
---------------------------------------------------------------------------
```

Table 10f. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
STREET NAME
STREET NUMBER
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 6				
ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.999
0.50%	0.9752	0.10000000E 01	0.10000000E 01	0.999
1.00%	0.9510	0.10000000E 01	0.10000000E 01	0.999
1.50%	0.9272	0.10000000E 01	0.10000000E 01	0.999
2.00%	0.9039	0.10000000E 01	0.10000000E 01	0.999
2.50%	0.8811	0.10000000E 01	0.10000000E 01	0.999
3.00%	0.8587	0.10000000E 01	0.10000000E 01	0.999
3.50%	0.8368	0.10000000E 01	0.10000000E 01	0.999
4.00%	0.8154	0.10000000E 01	0.10000000E 01	0.999
4.50%	0.7944	0.10000000E 01	0.10000000E 01	0.999
5.00%	0.7738	0.10000000E 01	0.10000000E 01	0.999
5.50%	0.7536	0.10000000E 01	0.10000000E 01	0.999
6.00%	0.7339	0.10000000E 01	0.10000000E 01	0.999

ble 10f. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
STREET NAME
STREET NUMBER
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 6

RECALL FACTOR	FILE SIZE		
	1 MILLION	4 MILLION	220 MILLION
1.0000	0.10000000E 01	0.10000000E 01	0.99999999E 00
0.9752	0.10000000E 01	0.10000000E 01	0.99999999E 00
0.9510	0.10000000E 01	0.10000000E 01	0.99999999E 00
0.9272	0.10000000E 01	0.10000000E 01	0.99999999E 00
0.9039	0.10000000E 01	0.10000000E 01	0.99999999E 00
0.8811	0.10000000E 01	0.10000000E 01	0.99999999E 00
0.8587	0.10000000E 01	0.10000000E 01	0.99999999E 00
0.8368	0.10000000E 01	0.10000000E 01	0.99999999E 00
0.8154	0.10000000E 01	0.10000000E 01	0.99999999E 00
0.7944	0.10000000E 01	0.10000000E 01	0.99999999E 00
0.7738	0.10000000E 01	0.10000000E 01	0.99999999E 00
0.7536	0.10000000E 01	0.10000000E 01	0.99999999E 00
0.7339	0.10000000E 01	0.10000000E 01	0.99999999E 00

Table 10g. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
STREET NAME
STREET NUMBER
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 7

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 M
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.9993
0.50%	0.9752	0.10000000E 01	0.10000000E 01	0.9993
1.00%	0.9510	0.10000000E 01	0.10000000E 01	0.9999
1.50%	0.9272	0.10000000E 01	0.10000000E 01	0.9999
2.00%	0.9039	0.10000000E 01	0.10000000E 01	0.9999
2.50%	0.8811	0.10000000E 01	0.10000000E 01	0.9999
3.00%	0.8587	0.10000000E 01	0.10000000E 01	0.9993
3.50%	0.8368	0.10000000E 01	0.10000000E 01	0.9999
4.00%	0.8154	0.10000000E 01	0.10000000E 01	0.9999
4.50%	0.7944	0.10000000E 01	0.10000000E 01	0.9999
5.00%	0.7738	0.10000000E 01	0.10000000E 01	0.9999
5.50%	0.7536	0.10000000E 01	0.10000000E 01	0.9999
6.00%	0.7339	0.10000000E 01	0.10000000E 01	0.9999

```
CISION OF RETRIEVED RESULTS FOR DATA FIELDS
   SEX
   STREET NAME
   STREET NUMBER
   DATE OF BIRTH (YEAR, MONTH, AND DAY)
   AND SURNAME
-----------------------------------------------------------------------------
                                                                            !
           SURNAME DECILE  7                                                !
-----------------------------------------------------------------------------
 !                              FILE SIZE                                   !
 !---------------------------------------------------------------------------
 !        1 MILLION        !         4 MILLION          !    220 MILLION     !
-----------------------------------------------------------------------------
 !                         !                            !                    !
 !     0.10000000E 01      !      0.10000000E 01        !   0.99999998E 00   !
 !                         !                            !                    !
 !     0.10000000E 01      !      0.10000000E 01        !   0.99999998E 00   !
 !                         !                            !                    !
 !     0.10000000E 01      !      0.10000000E 01        !   0.99999998E 00   !
 !                         !                            !                    !
 !     0.10000000E 01      !      0.10000000E 01        !   0.99999998E 00   !
 !                         !                            !                    !
 !     0.10000000E 01      !      0.10000000E 01        !   0.99999998E 00   !
 !                         !                            !                    !
 !     0.10000000E 01      !      0.10000000E 01        !   0.99999998E 00   !
 !                         !                            !                    !
 !     0.10000000E 01      !      0.10000000E 01        !   0.99999998E 00   !
 !                         !                            !                    !
 !     0.10000000E 01      !      0.10000000E 01        !   0.99999998E 00   !
 !                         !                            !                    !
 !     0.10000000E 01      !      0.10000000E 01        !   0.99999998E 00   !
 !                         !                            !                    !
 !     0.10000000E 01      !      0.10000000E 01        !   0.99999998E 00   !
 !                         !                            !                    !
 !     0.10000000E 01      !      0.10000000E 01        !   0.99999998E 00   !
 !                         !                            !                    !
 !     0.10000000E 01      !      0.10000000E 01        !   0.99999998E 00   !
 !                         !                            !                    !
 !     0.10000000E 01      !      0.10000000E 01        !   0.99999998E 00   !
-----------------------------------------------------------------------------
```

Table 10h. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
STREET NAME
STREET NUMBER
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 8

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MILL
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.9999999
0.50%	0.9752	0.10000000E 01	0.10000000E 01	0.9999999
1.00%	0.9510	0.10000000E 01	0.10000000E 01	0.9999999
1.50%	0.9272	0.10000000E 01	0.10000000E 01	0.9999999
2.00%	0.9039	0.10000000E 01	0.10000000E 01	0.9999999
2.50%	0.8811	0.10000000E 01	0.10000000E 01	0.9999999
3.00%	0.8587	0.10000000E 01	0.10000000E 01	0.9999999
3.50%	0.8368	0.10000000E 01	0.10000000E 01	0.9999999
4.00%	0.8154	0.10000000E 01	0.10000000E 01	0.9999999
4.50%	0.7944	0.10000000E 01	0.10000000E 01	0.9999999
5.00%	0.7738	0.10000000E 01	0.10000000E 01	0.9999999
5.50%	0.7536	0.10000000E 01	0.10000000E 01	0.9999999
6.00%	0.7339	0.10000000E 01	0.10000000E 01	0.9999999

```
ISION OF RETRIEVED RESULTS FOR DATA FIELDS
 SEX
 STREET NAME
 STREET NUMBER
 DATE OF BIRTH (YEAR, MONTH, AND DAY)
 AND SURNAME
-----------------------------------------------------------------------------
                                                                            !
         SURNAME DECILE  8                                                  !
----------------------------------------------------------------------------!
!                            FILE SIZE                                      !
!---------------------------------------------------------------------------!
!       1 MILLION        !         4 MILLION         !     220 MILLION      !
----------------------------------------------------------------------------!
!                        !                           !                      !
!     0.10000000E 01     !      0.10000000E 01       !   0.99999994E 00     !
!                        !                           !                      !
!     0.10000000E 01     !      0.10000000E 01       !   0.99999994E 00     !
!                        !                           !                      !
!     0.10000000E 01     !      0.10000000E 01       !   0.99999994E 00     !
!                        !                           !                      !
!     0.10000000E 01     !      0.10000000E 01       !   0.99999994E 00     !
!                        !                           !                      !
!     0.10000000E 01     !      0.10000000E 01       !   0.99999993E 00     !
!                        !                           !                      !
!     0.10000000E 01     !      0.10000000E 01       !   0.99999993E 00     !
!                        !                           !                      !
!     0.10000000E 01     !      0.10000000E 01       !   0.99999993E 00     !
!                        !                           !                      !
!     0.10000000E 01     !      0.10000000E 01       !   0.99999993E 00     !
!                        !                           !                      !
!     0.10000000E 01     !      0.10000000E 01       !   0.99999993E 00     !
!                        !                           !                      !
!     0.10000000E 01     !      0.10000000E 01       !   0.99999993E 00     !
!                        !                           !                      !
!     0.10000000E 01     !      0.10000000E 01       !   0.99999992E 00     !
!                        !                           !                      !
!     0.10000000E 01     !      0.10000000E 01       !   0.99999992E 00     !
!                        !                           !                      !
!     0.10000000E 01     !      0.10000000E 01       !   0.99999992E 00     !
-----------------------------------------------------------------------------
```

Table 10i. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
STREET NAME
STREET NUMBER
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 9

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	220 MILL
0.00%	1.0000	0.10000000E 01	0.10000000E 01	0.9999999
0.50%	0.9752	0.10000000E 01	0.10000000E 01	0.9999999
1.00%	0.9516	0.10000000E 01	0.10000000E 01	0.9999999
1.50%	0.9272	0.10000000E 01	0.10000000E 01	0.9999999
2.00%	0.9039	0.10000000E 01	0.10000000E 01	0.9999998
2.50%	0.8811	0.10000000E 01	0.10000000E 01	0.9999999
3.00%	0.8587	0.10000000E 01	0.10000000E 01	0.9999998
3.50%	0.8368	0.10000000E 01	0.10000000E 01	0.9999998
4.00%	0.8154	0.10000000E 01	0.10000000E 01	0.9999998
4.50%	0.7944	0.10000000E 01	0.10000000E 01	0.9999998
5.00%	0.7738	0.10000000E 01	0.10000000E 01	0.9999998
5.50%	0.7536	0.10000000E 01	0.10000000E 01	0.9999998
6.00%	0.7339	0.10000000E 01	0.10000000E 01	0.9999998

```
SION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
STREET NAME
STREET NUMBER
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME
---------------------------------------------------------------------------------------
                                                                                      !
       SURNAME DECILE  9                                                              !
--------------------------------------------------------------------------------------!
                                   FILE SIZE                                          !
--------------------------------------------------------------------------------------!
       1 MILLION           !              4 MILLION          !        220 MILLION     !
--------------------------------------------------------------------------------------!
                           !                                 !                        !
    0.10000000E 01         !          0.10000000E 01         !     0.99999990E 00     !
    0.10000000E 01         !          0.10000000E 01         !     0.99999990E 00     !
    0.10000000E 01         !          0.10000000E 01         !     0.99999990E 00     !
    0.10000000E 01         !          0.10000000E 01         !     0.99999990E 00     !
    0.10000000E 01         !          0.10000000E 01         !     0.99999989E 00     !
    0.10000000E 01         !          0.10000000E 01         !     0.99999989E 00     !
    0.10000000E 01         !          0.10000000E 01         !     0.99999989E 00     !
    0.10000000E 01         !          0.10000000E 01         !     0.99999989E 00     !
    0.10000000E 01         !          0.10000000E 01         !     0.99999988E 00     !
    0.10000000E 01         !          0.10000000E 01         !     0.99999988E 00     !
    0.10000000E 01         !          0.10000000E 01         !     0.99999988E 00     !
    0.10000000E 01         !          0.10000000E 01         !     0.99999987E 00     !
    0.10000000E 01         !          0.10000000E 01         !     0.99999987E 00     !
---------------------------------------------------------------------------------------
```

Table 10j. PRECISION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
STREET NAME
STREET NUMBER
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME

SURNAME DECILE 10

ERROR RATE	RECALL FACTOR	FILE SIZE		
		1 MILLION	4 MILLION	200 MILL
0.00%	1.0000	0.10000000E 01	0.99999999E 00	0.99999959
0.50%	0.9752	0.10000000E 01	0.99999999E 00	0.99999958
1.00%	0.9510	0.10000000E 01	0.99999999E 00	0.99999957
1.50%	0.9272	0.10000000E 01	0.99999999E 00	0.99999955
2.00%	0.9039	0.10000000E 01	0.99999999E 00	0.99999954
2.50%	0.8811	0.10000000E 01	0.99999999E 00	0.99999953
3.00%	0.8587	0.10000000E 01	0.99999999E 00	0.99999952
3.50%	0.8368	0.10000000E 01	0.99999999E 00	0.9999995
4.00%	0.8154	0.10000000E 01	0.99999999E 00	0.99999949
4.50%	0.7944	0.10000000E 01	0.99999999E 00	0.99999948
5.00%	0.7738	0.10000000E 01	0.99999999E 00	0.99999947
5.50%	0.7536	0.10000000E 01	0.99999999E 00	0.99999945
6.00%	0.7339	0.10000000E 01	0.99999999E 00	0.99999944

```
SION OF RETRIEVED RESULTS FOR DATA FIELDS
SEX
STREET NAME
STREET NUMBER
DATE OF BIRTH (YEAR, MONTH, AND DAY)
AND SURNAME
---------------------------------------------------------------------

        SURNAME DECILE 10
---------------------------------------------------------------------
                          FILE SIZE
---------------------------------------------------------------------
        1 MILLION        !        4 MILLION        !     220 MILLION
---------------------------------------------------------------------
    0.10000000E 01       !                         !
                         !    0.99999999E 00       !
    0.10000000E 01       !                         !  0.99999959E 00
                         !    0.99999999E 00       !
    0.10000000E 01       !                         !  0.99999958E 00
                         !    0.99999999E 00       !
    0.10000000E 01       !                         !  0.99999957E 00
                         !    0.99999999E 00       !
    0.10000000E 01       !                         !  0.99999955E 00
                         !    0.99999999E 00       !
    0.10000000E 01       !                         !  0.99999954E 00
                         !    0.99999999E 00       !
    0.10000000E 01       !                         !  0.99999953E 00
                         !    0.99999999E 00       !
    0.10000000E 01       !                         !  0.99999952E 00
                         !    0.99999999E 00       !
    0.10000000E 01       !                         !  0.99999951E 00
                         !    0.99999999E 00       !
    0.10000000E 01       !                         !  0.99999949E 00
                         !    0.99999999E 00       !
    0.10000000E 01       !                         !  0.99999948E 00
                         !    0.99999999E 00       !
    0.10000000E 01       !                         !  0.99999947E 00
                         !    0.99999999E 00       !
    0.10000000E 01       !                         !  0.99999945E 00
                         !    0.99999999E 00       !
                         !                         !  0.99999944E 00
---------------------------------------------------------------------
```

APPENDIX II

A PROBABILISTIC FORMULATION OF THE NON-UNIQUE ACCESS PROBLEM

In this appendix, we describe the theoretical basis of the probability model discussed in Section 5. The problem is to calculate the probability that a given retrieved record which has been accessed by a combination of retrieval keys such as name, date of birth, etc., will indeed correspond to an individual whose record is requested. It is to be noted that this theoretical development is completely independent of the retrieval algorithm itself. It supposes that a certain record has been retrieved and that it has certain characteristics which match characteristics known by the inquirer. The question is then: what is the probability that it is in fact the record desired by the inquirer?

Note that there is some conceptual difficulty in even defining precisely when a record is, in fact, the record of a particular person. This is because the record may not uniquely characterize a person or may contain erroneous information. There are also attributive values which carry more weight than others in identifying a record-person correspondence (e.g., name). For purposes of the discussion, however, we assume that it is meaningful to say that a record matches or corresponds to an individual, i.e., is the record of the individual. This proposition or event we designate as 'M_i'; that is,

M_i = a given record matches the individual i (1)

(e.g., who has attributes name = x_1, address = x_2, ..., $\text{attribute}_t = x_t$)

The second event to be specified in the probabilistic formulation is that a record contains a particular attribute value for the j^{th} retrieval key:

K_j = a given record has a particular attribute value for key j (e.g., name = x_1) (2)

It is useful to define a third "reference" event that defines the universe of discourse or context of the probability formulation. We call this A.

In what follows we can consider A to be the selection of a given record from a data bank where the mode of selection is not specified.

The basic probability to be calculated is then

$$P(M_i | K_1 \ \& \ K_2 \ \& \ \dots \ \& \ K_n \ \& \ A) \tag{3}$$

where '&' indicates logical conjunction. Thus (3) is the probability that a given record matches the individual i conditional on the fact that the record contains certain attribute values on the set of keys $j = 1, \ldots, n$. The basic problem of non-unique access is to select a set of keys so that the probability (3) is near unity.

Let us first solve the problem of calculating probability (3) by studying the case of a single key, i.e., the probability

$$P(M_i | K_j \ \& \ A) \tag{4}$$

By the multiplication principle we have

$$P(M_i | K_j \ \& \ A) = \frac{P(M_i | A) \ P(K_j | M_i \ \& \ A)}{P(K_j | A)} \tag{5}$$

For readability we will omit the reference event A, thus (5) assumes the simpler form.

$$P(M_i | K_j) = \frac{P(M_i) \ P(K_j | M_i)}{P(K_j)} \tag{6}$$

The factors on the right have the following meaning:

$$P(M_i) = \text{the a priori probability that a randomly selected record matches the individual } i \tag{7}$$

$P(M_i) = 1/N$ if N is the total population upon which the data bank is drawn.[*] For example, if the individual i is known to be a California driver and the data bank contains all and only California driver records, then N = the size of the data bank. On the other hand, if the data bank concerns information on adult U.S. citizens, then N = the number of adult U.S. citizens.

---

[*] In the complete formulation with the reference event A, A could be considered to contain the information regarding the population upon which the data bank is based.

We next have:

$P(K_j|M_i)$ = the probability that a record contains a specific value for retrieval key j conditional on the fact that the record matches the individual i. (8a)

For example, if the j^{th} key is i's address, then $P(K_j|M_i)$ is the probability that i's record contains his correct address. It therefore follows that

$P(K_j|M_i)$ = the complement of the sum of the error rate in key j and the relative frequency of omission of data in key j (8b)

Note that for perfectly accurate records, we would have $P(K_j|M_i) = 1$.*

The third factor $P(K_j)$ is estimated as the relative frequency of the specific information given by key j in the bank. For example, if two records are retrieved having the same value for key j and m is the size of the data bank then $P(K_j)$ is estimated as 2/m.

Now if a set of keys j = 1, ..., n are independent, the probability (3) is calculated by putting the conjunction $K_1 \& \ldots \& K_n$ for K_j in (5) (or the abbreviated Form (6)) and expanding the result into a product for j = 1, ..., n; i.e.,

$$P(M_i|K_1\& \ldots \&K_n) = P(M_i) \prod_{j=1}^{n} \frac{P(K_j|M_i)}{P(K_j)} \tag{9}$$

However, the independence assumption is only valid for certain keys. For example, name and birthdate can be regarded as independent, but not birthdate and year of graduation from high school. Independence also depends on the conditional probability. We therefore must look for a more general solution.

---

* The error rate in query formulation is not part of the probability (8), which has to do with the probability of the selection procedure being successful conditional on an error-free query.

One approach to this problem is to use some probability machinery developed by W. E. Johnson and reported in Keynes.[53] There are also versions of the Johnson work in Carnap.[54]

We now present our version of the so-called "Johnson Cumulative Formula." A set of relevance coefficients or coefficients of influence is first defined. For any propositions (events) $K_1, \ldots, K_n$, we define

$$C(K_1, K_2|A) = \frac{P(K_1|K_2 \& A)}{P(K_1|A)} \tag{10}$$

$$C(K_1, K_2 \ldots, K_n|A) = C(K_1, K_2|A) \cdot C(K_1 \& K_2, K_3, \ldots, K_n|A) \tag{11}$$

It is then easy to show that $C(K_1, \ldots, K_n|A)$ has the operator property*

$$P(K_1 \& \ldots \& K_n|A) = C(K_1, \ldots, K_n|A) \prod_{j=1}^{n} P(K_j|A) \tag{12}$$

(see Carnap[55]).

The coefficient (11) is thus a coefficient of independence and is unity if the events $K_1, K_2, \ldots, K_n$ are independent relative to the event A.

The cumulative formula is obtained as follows:

Let K be the conjunction $K_1 \& \ldots \& K_n$. Then by the multiplication principle

$$P(M_i|K \& A) = \frac{P(M_i|A) \quad P(K|M_i \& A)}{P(K|A)} \tag{13}$$

Applying (12) to the numerator and denominator of (13) and then multiplying both sides of the result by $[P(M_i|A)]^{n-1}$ we get

---

* We use our own notation for the coefficient.

[53] Keynes, J. M., A Treatise on Probability, MacMillan and Co., London, 1929, pp. 149-155.

[54] Carnap, R., The Logical Foundations of Probability, University of Chicago Press, 1950, pp. 356-360.

[55] Ibid, p. 358, Formula T66-3C.

$$[P(M_i|A)]^{n-1} \cdot P(M_i|K \,\&\, A) = \frac{C(K_1, \ldots, K_n|M_i \,\&\, A)}{C(K_1, \ldots, K_n|A)} \prod_{j=1}^{n} \left(\frac{P(M_i|A)\; P(K_j|M_i \,\&\, A)}{P(K_j|A)} \right) \tag{14}$$

We now use the cumulative formula (14) in the following way. First introduce the abbreviations.

$$V = P(M_i|K \,\&\, A) \tag{15}$$

$$p = P(M_i|A) \tag{16}$$

$$q_j = P(M_i|K_j \,\&\, A) \tag{17}$$

By the multiplication principle, q_j is the factor in the parentheses of (14), i.e.,

$$q_j = \frac{P(M_i|A)\; P(K_j|M_i \,\&\, A)}{P(K_j|A)} \tag{18}$$

Our problem is to study the expression for V. Using the abbreviations, the formula (14) becomes

$$p^{n-1}V = \frac{C(K_1, \ldots, K_n|M_i \,\&\, A)}{C(K_1, \ldots, K_n|A)} \prod_{j=1}^{n} q_j \tag{19}$$

We now use (19) again, but this time substitute '$\sim M_i$', i.e., the negation of M_i, for 'M_i' in (14). The quantities p, V, and q_j are then replaced by their complements. We get.

$$(1-p)^{n-1}\,(1-V) = \frac{C(K_1, \ldots, K_n|\sim M_i \,\&\, A)}{C(K_1, \ldots, K_n|A)} \prod_{j=1}^{n} (1 - q_j) \tag{20}$$

The ratio of (20) to (19) yields

$$\frac{1}{V} - 1 = E\left(\frac{p}{1-p}\right)^{n-1} \prod_{j=1}^{n} \left(\frac{1-q_j}{q_j}\right) \tag{21}$$

where E is the factor

$$E = \frac{C(K_1, \ldots, K_n | {\sim}M_i \ \& \ A)}{C(K_1, \ldots, K_n | M_i \ \& \ A)} \tag{22}$$

The equation (21) is the solution to our problem; that is, the right side of (21) is the product of the two quantities, E as defined by (22), and

$$F = \left(\frac{p}{1-p}\right)^{n-1} \prod_{j=1}^{n} \left(\frac{1 - q_j}{q_j}\right) \tag{23}$$

so that (21) becomes

$$\frac{1}{V} - 1 = EF$$

from which we derive

$$V = \frac{1}{1 + EF} \tag{24}$$

We see that V depends not only upon the quantities q_j but also on p. the a priori probability of M_i, and the quantity E. The value used for E is usually obtained by independence assumptions.

Note that if

$$E < 1$$

then, from (24),

$$V > \frac{1}{1 + F} \qquad (25)$$

This means that a lower limit to the probability of M_i conditional on the combination K_1 & K_2 & ... & K_n can be computed if $E \leq 1$. The value $E = 1$ occurs when the numerator and denominator of (22) are equal. In particular, this is the case when the keys are independent relative to both $\sim M_i$ and M_i. The smaller the value of E, the greater the ability of the keys to discriminate.

To summarize: We have three formulas for the desired probability V each involving a different ratio of coefficients of influence: the formula (19), the formula (20), and the formula (24). In Section 5 we have implemented the formula (24) for a range of values of E.

APPENDIX III

SOFT-MATCH TECHNIQUES

In Section 5.2, we noted the classical problem of trade-off between recall and precision, i.e., increasing recall lowers precision. But this arises only in the case where all retrieved records are treated on an equal par and not subjected to further filtering. Our approach to the problem is to increase the recall and then to provide a mechanism for additional filtering to improve the precision. In this section we discuss a method for increasing the recall called "Soft-Matching." The result of a Soft-Match Technique is not only a set of retrieved records, but also a numerical weight assigned to each record to indicate its closeness to the query. As an example, consider the case of a query which is a conjunction of three key values denoted by A, B, C. A traditional retrieval would select only records that matched on all three. A soft-match technique might be to also select those records that matched on A and B but not on C, on A and C but not on B, and on B and C but not on A. These secondary retrievals would then be assigned weights less than that assigned to a record that matched on all three. The retrieval could also be extended to tertiary retrievals such as records that matched on A but not on B and not on C, etc. These would be given the lowest weights.

In this appendix a heuristic weighing scheme is presented for filtering the candidate records. In Section 5.5 and Appendix II we presented a second weighting scheme based on probabilities which is in fact the measure of precision for a single retrieval. (That is to say, precision is the proportion of retrieved documents that are relevent, i.e., match the inquirer. Consequently, if exactly one record is retrieved, then the precision corresponds exactly to the probability that the record is in fact that desired by the inquirer.)

On the basis of the above discussion, we see that a soft-match technique involves two components: (1) an algorithm for determining a subset of the data base of manageable size which has maximum recall; (2) a scoring function for assigning the closeness measures to each item in the subset.

First, consider the problem of determining a manageable subset. Suppose that the retrieval prescription involves n keys

$$K_1, K_2, \ldots, K_n$$

Then let $S(K_i)$ be the set of records which match on the i^{th} key. In principle, the subset with maximum recall would be the set union of $S(K_1)$, ..., $S(K_n)$. This approach would not be efficient however, so a method which utilizes the error and omission rates and the discrimination factors is recommended. These quantities are defined as follows:

a_i = probability that the key value for the i^{th} key is incorrect

b_i = probability that the key value for the i^{th} key is omitted

$e_i = a_i + b_i$

(Thus, the complement of e_i is the quantity c_i defined in Section 5.1). The discrimination factor of a key is measured by the relative frequency of a key value in the data base. Let R_i be the relative frequency for the i^{th} key. This quantity R_i is estimated for certain keys by the values given in Tables 5.3-1, 5.3-2; it therefore corresponds to the concept of retrieval ratio.

It follows that the initial selection of the retrieval should be the union of, not all n sets $S(K_i)$, but a certain selection of fewer sets say $S(K_{i_1})$, ..., $S(K_{i_m})$ $(m<n)$, that has the properties

(1) $R_{i_1} \times R_{i_2} \times \ldots \times R_{i_m}$ is less than a certain threshold r_o.

(2) $e_{i_1} \times e_{i_2} \times \ldots \times e_{i_m}$ is less than a certain threshold e_o.

The first property guarantees that the retrieval set has manageable size. The second property guarantees that the probability of missing a relevant record is low.

The second aspect of the soft-match technique requires the definition of the scoring function. For this purpose consider that we have an arbitrary quantitative "strength of match" function for each key. For example, for a key designating home address, one such function might be

$f_i = 1$ for complete match

$f_i = .5$ for agreement on the street name, part of the home address but not on the house number part

$f_i = 0$ otherwise.

More complicated functions can be defined in obvious ways. For example, if the house number partially matched, say it disagreed on one numeral out of four, then we could score this as .75. The simplest function is the binary function, '1' for match, '0' for no match.

In general, we see that a score for a key for which $f_i > 0$ (i.e., a partially matching key) should:

(1) Increase as R_i decreases

(2) Increase as f_i increases

(3) Increase as a_i (the error rate in the i^{th} key) decreases.

It is therefore plausible to take the contribution of the i^{th} key to the total score (in the case that $f_i > 0$) to be proportional to the quantity

$$\frac{f_i\ (1 - a_i)}{R_i}$$

since this is a simple expression that satisfies the requirements (1), (2), and (3).

It is convenient to normalize the scores, however, so we will take the constant of proportionality to be 1/R where

$$R = \sum_{i=1}^{n} \frac{1}{R_i} \tag{1}$$

Thus, the score for the i^{th} key is

$$S_i = \frac{1}{R}\ \frac{f_i\ (1-a_i)}{R_i} \tag{2}$$

We now define the total score as consisting of two components: a positive component made up of the contributions S_i defined above and a negative component to be defined below.

The positive component we take to be simply

$$S = \sum_{i=1}^{n} S_i \tag{3}$$

(Note that this summation runs over all the keys in the query, and not over just the subset of m keys used to get maximum recall.)

By (1) and (2) we see that S is a convex linear combination of the terms $f_i(1-a_i)$ where i = 1, ..., n, and so S can range from 0 to 1.

To define the negative component, we see that if a key is error free and fails to match ($f_i = 0$), then it should kill off the entire retrieval. Thus, in the case $f_i = 0$ we should have a negative contribution which varies as $(1 - a_i)$. We therefore take the score when $f_i = 0$ to be

$$T_i = -(1 - a_i)$$

Since a value of $T_i = -1$ for only one key is sufficient to invalidate a retrieval, we take the total negative score to be, not the sum of the individual scores, but rather the negative of the maximum value of $(1 - a_i)$, i.e.,

$$T = -\max_{f_i=0} (1 - a_i)$$

Thus T ranges from -1 to 0.

The retrieval scoring function is thus

$$U = S + T$$

An example of how this scoring function is used is given in Table III-1. The example deals with the scoring of four records based on values of the parameters R_i (the retrieval ratio) and a_i (the error rate).

Table III-1. Example of Soft-Match

	KEY		
	NAME	ADDRESS	SEX
R_i	.003	.00002	.5
a_i	.001	.01	0

Table of Values for R_i, a_i

	NAME	ADDRESS	SEX
Record 1	match (1)	match (1)	no match (0)
Record 2	match (1)	partial match (.75)*	match (1)
Record 3	partial match (.83)**	match (1)	match (1)
Record 4	match (1)	match (1)	data absent

Table of Retrieval Outcomes
(values in parentheses are those for f_i)

	S	T	U
Record 1	.9900	-1	-.0100
Record 2	.7442	0	.7442
Record 3	.9889	0	.9889
Record 4	.9900	0	.9900

Retrieval Scores

* The partial match example is for the case that one character out of four doesn't match.

** The partial match example is for the case that one character out of six doesn't match.

APPENDIX IV

FORTRAN SOURCE CODE FOR THE PROBABILITY MODEL

```
FORTRAN IV      D06-01  SOURCE LISTING                  PAGE 001

      C**********************
      C   QUERY SIMULATES THE DATA BASE/USER INTERACTION
      C**********************
            DIMENSION A(11).B(11).C(11).R(11).S(11)
            DIMENSION JARY(11).IFLD(11)
            DIMENSION ARGS(11)
            DIMENSION SURNM(10)
            COMMON IRAN. JRAN. EARY(27).PARY(3)
            DATA ARGS(1)/.239726E-2/.ARGS(2)/.833333E-1/.ARGS(3)/.02/
            DATA ARGS(4)/.02/.ARGS(5)/.5/.ARGS(6)/.0001/.ARGS(7)/.0002/
            DATA ARGS(8)/.0666666/.ARGS(9)/.000333333/.ARGS(10)/.555555E-4/
            DATA SURNM(1).SURNM(2).SURNM(3).SURNM(4)/.907E-7..2459E-5..9635E-5
           2 ..272037E-4/
            DATA SURNM(5).SURNM(6).SURNM(7)/4.81966E-5.6.627E-5.1.265E-4/
            DATA SURNM(8).SURNM(9).SURNM(10)/4.8197E-4.7.832E-4..3384E-2/
            DATA        PARY(1).PARY(2).PARY(3)/4.545454E-9.2.5E-7.1.E-6/
            DATA EARY(1).EARY(2).EARY(3).EARY(4)/.5.1..2..10./
      C**********************
      C
      C   QUERY SIMULATES QUERY PROCESSING AGAINST A DATA BASE.
      C   IT GENERATES VALUES FOR A(J)[THE ERROR RATE IN A FIELD]
      C   B(J)  [THE OMISSION RATE OF DATA IN A FIELD]
      C   C(J)  [1-(A(J)+B(J)]...THE CONFIDENCE FACTOR
      C   R(J)  THE PERCENTAGE OF THE FILE RETRIEVED VIA A FIELD. AND
      C   S(J)  A QUERY SCORE. WHICH RANGES FROM 1 TO THE FILE SIZE.
      C   S(J)  IS NOMINALLY THE RATIO C(J)/R(J). AND IS ADJUSTED HERE
      C   TO MAINTAIN REASONABLE LIMITS.
      C   THE USER ENTERS THE NUMBER OF ARGUMENTS (FROM 1 TO 11)
      C   WHICH HE WISHES TO USE IN HIS QUERY.  THE QUERY IS TREATED AS IF
      C   IT REQUIRED A SUCCESSFUL BOOLEAN AND OF ALL THE ARGUMENTS TO BE
      C   SUCCESSFUL.
      C   THE PROGRAM WILL TERMINATE WHEN THE NUMBER OF INPUT ARGUMENTS
      C   IS SET TO ZERO.
      C
      C   FOR THIS TEST THE VALUES OF A(J) AND B(J) WILL BE TREATED AS ZERO.
      C   FOLLOWING THE ENTRY OF THE FIRST NUMBER--THE NUMBER OF ARGUMENTS
      C   TO FOLLOW--THE USER MAY ENTER NUMBERS CORRESPONDING TO THE ACTUAL
      C   ARGUMENTS.  EACH ARGUMENT HAS AN ASSOCIATED R(J). WHICH IS THE
      C   PERCENTAGE RETRIEVAL TO BE EXPECTED.  THE ARGUMENT NUMBERS AND
      C   THEIR ASSOCIATED VALUES ARE AS FOLLOWS:
      C
      C   ARGUMENT NUMBER       MEANING                   RETRIEVAL RATIO
      C
      C    1             DATE OF BIRTH (MMDD)     1/365
      C    2             MONTH OF BIRTH           1/12
      C    3             YEAR OF BIRTH            1/50
      C    4             STATE                    1/50
      C    5             SEX                      1/2
      C    6             STREET NUMBER            1/10.000
      C    7             STREET NAME              1/5.000
      C    8             EDUCATION                1/15
      C    9             COUNTY                   1/3000 (ESTIMATED)
      C          10            DATE OF BIRTH (YMD)      1/10.000
      C**********************
      C   BEGIN BY ASKING FOR OPERATOR INPUT.
      C**********************
      1     WRITE(6.205)
      C**********************
      C   NOW EXAMINE NARGS TO START THINGS GOING.
      C*********************
            READ (6,210) NARGS
            IF (NARGS) 100, 100, 10
      C*********************
      C   COME HERE TO PREPARE A QUERY SURROGATE.
      C*********************
      10    DO 20 I=1,NARGS
            WRITE      (6,220)
            READ      (6,210) J
            A(I)=0.
            B(I)=0.
            C(I)=1.
            R(I)=ARGS(J)
            S(I)=(C(I)/R(I))
            JARY(I)=J
      20    CONTINUE
            WRITE (6,215)
      C*********************
      C   NOW CALL ON THE MAIN COMPUTATIONAL ROUTINE TO PREPARE THE TABLES.
      C*********************
            JARY(NARGS+1)=11
            N=NARGS+1
            WRITE(5,200) NARGS
            DO 25 J=1,NARGS
            WRITE(5,245) J,A(J),J,B(J),J,C(J),J,R(J),J,S(J),J,JARY(J)
      25    CONTINUE
            DO 50 I=1,10
            C(NARGS+1)=1.
            B(NARGS+1)=0.
            A(NARGS+1)=3.
            R(NARGS+1)=SURNM(I)
            S(NARGS+1)=C(NARGS+1)/R(NARGS+1)
            WRITE(5,250)
            DO 45 J=1,NARGS
            GOTO (31,32,33,34,35,36,37,38,39,40) JARY(J)
      31    WRITE(5,255)
            GOTO 45
      32    WRITE(5,260)
            GOTO 45
      33    WRITE(5,265)
            GOTO 45
      34    WRITE(5,270)
            GOTO 45
      35    WRITE(5,275)
            GOTO 45
      36    WRITE(5,280)
            GOTO 45
      37    WRITE(5,285)
            GOTO 45
      38    WRITE(5,290)
            GOTO 45
      39    WRITE(5,295)
            GOTO 45
      40    WRITE(5,300)
      45    CONTINUE
            WRITE(5,305)
            WRITE(5,150)
            WRITE (5,225) I
            WRITE(5,230)
```

```
           WRITE(5.235)
           WRITE(5.240)
           CALL PVCY(A.B.C.R.S.N.JARY)
      50   CONTINUE
           GO TO 1
      100  CALL PVCY(A.B.C.R.S.NARGS.JARY)
           CALL EXIT
      C*********************
      C    FORMAT STATEMENTS
      C*********************
      150  FORMAT(1H .13X.110('-').7X)
      200  FORMAT(1H1.'NARGS='.I2)
      205  FORMAT ('0   ENTER NUMBER OF ARGUMENTS'/)
      210  FORMAT (I2)
      215  FORMAT (' EXITING QUERY')
      220  FORMAT (' ENTER QUERY FIELD IDENTIFIER')
      225  FORMAT(1H .13X.'!'.108X.'!'.7X/1H .13X.'!'.43X.
          1 'SURNAME DECILE'.1X.I2.48X.'!'.7X)
      230  FORMAT(1H .13X.'!'.108('-').'!'.7X/
          1 1H .13X.'!'.3X.'ERROR RATE'.2X.'!'.2X.'RECALL FACTOR'.
          2 2X.'!'.28X.'FILE SIZE'.37X.'!'.7X)
      235  FORMAT(1H .13X.'!'.15X.'!'.17X.'!'.74('-').'!'.7X/
          1 1H .13X.'!'.15X.'!'.17X.'!'.7X.'1 MILLION'.7X.'!'.10X.
          2 '4 MILLION'.8X.'!'.5X.'220 MILLION'.6X.'!'.7X)
      240  FORMAT(1H .13X.'!'.108('-').'!'.7X)
      245  FORMAT(1H0.'A('.I2.')='.E15.8.3X.'B('.I2.')='.E15.8.
          1 3X.'C('.I2.')='.E15.8.3X.'R('.I2.')='.E15.8.3X.'S('.
          2 I2.')='.E15.8.3X.'JARY('.I2.')='.I2)
      250  FORMAT(1H1.43X.'PRECISION OF RETRIEVED RESULTS FOR DATA
          1 FIELDS')
      255  FORMAT(1H .48X.'DATE OF BIRTH (MONTH AND DAY ONLY)')
      260  FORMAT(1H .48X.'MONTH OF BIRTH')
      265  FORMAT(1H .48X.'YEAR OF BIRTH')
      270  FORMAT(1H .48X.'STATE')
      275  FORMAT(1H .48X.'SEX')
      280  FORMAT(1H .48X.'STREET NUMBER')
      285  FORMAT(1H .48X.'STREET NAME')
      290  FORMAT(1H .48X.'EDUCATION')
      295  FORMAT(1H .48X.'COUNTY')
      300  FORMAT(1H .48X.'DATE OF BIRTH (YEAR. MONTH. AND DAY)')
      305  FORMAT(1H .48X.'AND SURNAME')
           END
```

```
FORTRAN IV      STORAGE MAP

A         000006  REAL*4     ARRAY (11)
B         000062  REAL*4     ARRAY (11)
C         000136  REAL*4     ARRAY (11)
R         000212  REAL*4     ARRAY (11)
S         000266  REAL*4     ARRAY (11)
JARY      000342  INTEGER*2  ARRAY (11)
IFLD      000370  INTEGER*2  ARRAY (11)
ARGS      000416  REAL*4     ARRAY (11)
SURNM     000472  REAL*4     ARRAY (10)
NARGS     002100  INTEGER*2  VARIABLE
I         002102  INTEGER*2  VARIABLE
J         002104  INTEGER*2  VARIABLE
N         002106  INTEGER*2  VARIABLE
PVCY      000000  REAL*4     PROCEDURE
EXIT      000000  REAL*4     PROCEDURE

COMMON BLOCK //           LENGTH 000174

IRAN      000000  INTEGER*2  VARIABLE
JRAN      000002  INTEGER*2  VARIABLE
EARY      000004  REAL*4     ARRAY (27)
PARY      000160  REAL*4     ARRAY (3)
```

```
FORTRAN IV      D06-01  SOURCE LISTING                          PAGE 001

      C*********************
      C       MAIN COMPUTATION ROUTINE OF PRIVACY MODEL--MARCH 1976
      C       **NOTE: THIS SUBROUTINE HAS BEEN MODIFIED TO PRODUCE A TABULAR PRINTOUT.
      C*********************
              SUBROUTINE PVCY(AJ.BJ.CJ.RJ.SJ.NARGS.JARY)
      C*********************
      C DECLARE VARIABLES
      C*********************
              DIMENSION JARY(11)
              DOUBLE PRECISION BIGPI(11).F.VARY(13.3)
              COMMON IRAN.JRAN.EARY(27),PARY(3)
              REAL AJ(11).BJ(11).CJ(11).RJ(11).SJ(11).P
              REAL QARY(11)
              DATA VARY(1.1)/9.0D0/
      C*********************
      C MAIN LOOP STARTS HERE
      C
      C VALUES ARE RETURNED IN THE ARGUMENT ARRAYS--TERMINATE IF NARGS< = 0
      C*********************
              IF(NARGS) 95.95.11
      C*********************
      C BEGIN PROCESSING THE QUERY RESPONSE HERE
      C*********************
       11     CONTINUE
              DO 90 JK=1.13
              L=JK
              CONF=1.-0.005*(L-1)
              REFAC=CONF**(NARGS)
              ERRAT=1.-CONF
              PERERR=ERRAT*100
              DO 80 I=1.3
              P=PARY(I)
              DO 20 II=1.NARGS
       20     QARY(II)=P*SJ(II)*CONF
      C*********************
      C NOW COMPUTE THE FUNCTION F
      C*********************
              F=1.
              X=P/(1.-P)
              DO 30 JJ=1.NARGS
              BIGPI(JJ)=(1.-QARY(JJ))/QARY(JJ)
       30     F=F*X**BIGPI(JJ)
              F=F/X
      C*********************
      C PREPARE THE PRINTOUT TABLE
      C*********************
              VARY(JK.I)=1./(1.+F)
              IF(VARY(JK.I).GT.(1.0)) VARY(JK.I)=1.0
       80     CONTINUE
              WRITE(5.150) PERERR.REFAC.(VARY(JK.I).I=3.1.-1)
       90     CONTINUE
              WRITE(5.170)
              RETURN
      C*********************
      C EXIT VIA HERE WHEN NARGS=0
      C*********************
       95     WRITE(5.200)
              RETURN
      C*********************
```

```
FORTRAN IV       D06-01  SOURCE LISTING                         PAGE 002

      C FORMAT STATEMENT SECTION
      C*********************************
0036  150   FORMAT(1H ,13X,'!',15X,'!',17X,'!',23X,'!',27X,'!',22X,'!',7X/
           1 1H ,13X,'!',4X,F6.2,'%',4X,'!',4X,F8.4,5X,'!',4X,
           2 F15.8,4X,'!',6X,E15.8,6X,'!',3X,E15.8,4X,'!',7X)
0037  170   FORMAT(1H ,13X,118('-'),7X)
0038  200   FORMAT(1H0,'    NORMAL TERMINATION')
0039        END
```

```
FORTRAN IV      STORAGE MAP

JARY     000030  INTEGER*2 PARAMETER ARRAY (11)
BIGPI    000032  REAL*8    ARRAY (11)
VARY     000162  REAL*8    ARRAY (13.3) VECTORED
AJ       000014  REAL*4    PARAMETER ARRAY (11)
BJ       000016  REAL*4    PARAMETER ARRAY (11)
CJ       000020  REAL*4    PARAMETER ARRAY (11)
RJ       000022  REAL*4    PARAMETER ARRAY (11)
SJ       000024  REAL*4    PARAMETER ARRAY (11)
QARY     000652  REAL*4    ARRAY (11)
NARGS    000026  INTEGER*2 PARAMETER VARIABLE
F        001140  REAL*8    VARIABLE
P        001150  REAL*4    VARIABLE
JK       001154  INTEGER*2 VARIABLE
L        001156  INTEGER*2 VARIABLE
CONF     001160  REAL*4    VARIABLE
REFAC    001164  REAL*4    VARIABLE
ERRAT    001170  REAL*4    VARIABLE
PERERR   001174  REAL*4    VARIABLE
I        001200  INTEGER*2 VARIABLE
II       001202  INTEGER*2 VARIABLE
X        001204  REAL*4    VARIABLE
JJ       001210  INTEGER*2 VARIABLE

COMMON BLOCK //          LENGTH 000174

IRAN     000000  INTEGER*2 VARIABLE
JRAN     000002  INTEGER*2 VARIABLE
EARY     000004  REAL*4    ARRAY (27)
PARY     000160  REAL*4    ARRAY (3)
```

NBS-114A (REV. 7-73)

U.S. DEPT. OF COMM. **BIBLIOGRAPHIC DATA SHEET**	1. PUBLICATION OR REPORT NO. NBS SP-500-2	2. Gov't Accession No.	3. Recipient's Accession No.

4. TITLE AND SUBTITLE

Accessing Individual Records From Personal
Data Files Using Non-Unique Identifiers

5. Publication Date
February 1977

6. Performing Organization Code

7. AUTHOR(S) Gwendolyn B. Moore, John L. Kuhns,
Jeffrey L. Trefftzs, Christine A. Montgomery

8. Performing Organ. Report No.

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Operating Systems, Inc.
21031 Ventura Boulevard
Woodland Hills, California 91364

10. Project/Task/Work Unit No.
640.1118

11. Contract/Grant No.
5-35928

12. Sponsoring Organization Name and Complete Address *(Street, City, State, ZIP)*

National Bureau of Standards
Department of Commerce
Washington, D.C. 20234

13. Type of Report & Period Covered
Final 31 March 1976

14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

Library of Congress Catalog Card Number: 76-57950

16. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

The Privacy Act of 1974 places restrictions on the Federal, state and local agencies' use of the Social Security account number as an identifier. For some agencies, compliance will involve changes in implementation of retrieval algorithms. This report describes methodology applicable to these changes in the more general context of the problem of retrieving individual records from files using non-unique identifiers. State-of-the-art retrieval techniques are discussed, a method for assigning reliability weights to various personal data elements is presented, file validation techniques for the error and omission rates of data items are suggested, and a retrieval probability model--designed to show likelihood of retrieval of a subject's record given a variety of populations, combinations of identifiers, and error/omission rates--is described. A methodology is developed for forming confidence factors from the established error/omission rates for combinations of non-unique identifiers that are candidates for use as retrieval keys. Use of these confidence factors as indices into the precision tables produced by the probability model is described.

17. KEY WORDS *(six to twelve entries; alphabetical order; capitalize only the first letter of the first key word unless a proper name; separated by semicolons)*

Data retrieval; file validation; name lookup; non-unique identifiers; personal data files; Privacy Act; probability model; retrieval.

18. AVAILABILITY ☒ Unlimited

☐ For Official Distribution. Do Not Release to NTIS

☒ Order From Sup. of Doc., U.S. Government Printing Office
Washington, D.C. 20402, SD Cat. No. C13.10:500-2

☐ Order From National Technical Information Service (NTIS)
Springfield, Virginia 22151

19. SECURITY CLASS (THIS REPORT)
UNCLASSIFIED

20. SECURITY CLASS (THIS PAGE)
UNCLASSIFIED

21. NO. OF PAGES
203

22. Price
$2.65

USCOMM-DC 29042-P74

U.S. GOVERNMENT PRINTING OFFICE: 1977 240-848/120 1-3

ANNOUNCEMENT OF NEW PUBLICATIONS ON COMPUTER SCIENCE & TECHNOLOGY

Superintendent of Documents,
Government Printing Office,
Washington, D. C. 20402

Dear Sir:

Please add my name to the announcement list of new publications to be issued in the series: National Bureau of Standards Special Publication 500-.

Name ____

Company ____

Address ____

City ____________________ State ______________ Zip Code __________

(Notification key N-503)